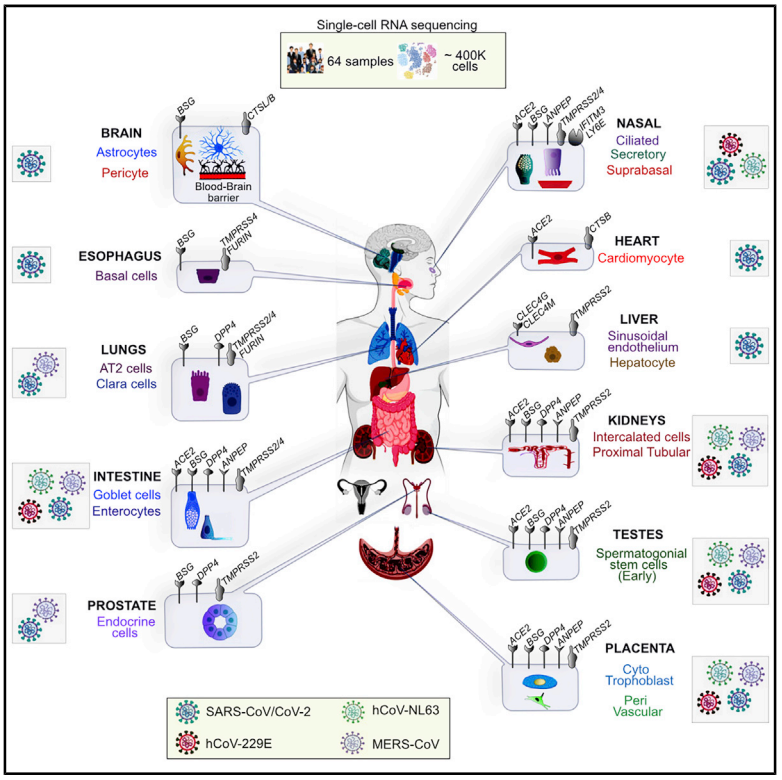
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cell Reports

Resource

A Single-Cell RNA Expression Map of Human Coronavirus Entry Factors

Graphical Abstract



Highlights

- Single-cell transcriptome profiling of SCARFs in various somatic and reproductive tissues
- Intestine, kidneys, placenta, and spermatogonia appear most permissive for coronavirus
- Nasal epithelium exhibits high expression of both promoting and restricting factors
- SCARF expression is conserved across the major organs of human, chimpanzee, and macaque

Authors

Manvendra Singh, Vikas Bansal, Cédric Feschotte

Correspondence

ms3559@cornell.edu (M.S.), vikas.bansal@dzne.de (V.B.), cf458@cornell.edu (C.F.)

In Brief

Singh et al. provide a resource to predict the tropism and identify the plausible entry points for SARS-CoV-2 and other pathogenic coronaviruses throughout the human body. The RNA levels of 28 genes dubbed "SCARFs," for SARS-CoV-2 and coronavirus-associated receptors and factors, are profiled in human tissues at the single-cell level.

Singh et al., 2020, Cell Reports *32*, 108175
September 22, 2020
https://doi.org/10.1016/j.celrep.2020.108175

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

A Single-Cell RNA Expression Map of Human Coronavirus Entry Factors

Manvendra Singh,[1,4,*] Vikas Bansal,[2,3,4,*] and Cédric Feschotte[1,5,*]
[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Biomedical Data Science and Machine Learning Group, German Center for Neurodegenerative Diseases, Tübingen 72076, Germany
[3]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen 37075, Germany
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: ms3559@cornell.edu (M.S.), vikas.bansal@dzne.de (V.B.), cf458@cornell.edu (C.F.)
https://doi.org/10.1016/j.celrep.2020.108175

SUMMARY

To predict the tropism of human coronaviruses, we profile 28 SARS-CoV-2 and coronavirus-associated receptors and factors (SCARFs) using single-cell transcriptomics across various healthy human tissues. SCARFs include cellular factors both facilitating and restricting viral entry. Intestinal goblet cells, enterocytes, and kidney proximal tubule cells appear highly permissive to SARS-CoV-2, consistent with clinical data. Our analysis also predicts non-canonical entry paths for lung and brain infections. Spermatogonial cells and prostate endocrine cells also appear to be permissive to SARS-CoV-2 infection, suggesting male-specific vulnerabilities. Both pro- and anti-viral factors are highly expressed within the nasal epithelium, with potential age-dependent variation, predicting an important battleground for coronavirus infection. Our analysis also suggests that early embryonic and placental development are at moderate risk of infection. Lastly, SCARF expression appears broadly conserved across a subset of primate organs examined. Our study establishes a resource for investigations of coronavirus biology and pathology.

INTRODUCTION

The zoonotic spillover of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the human population is causing a disease known as coronavirus disease 2019 (COVID-19) (Lu et al., 2020; Paules et al., 2020). Since the first case reported in late December 2019, SARS-CoV-2 has spread to 215 countries, infecting more than 20 million humans and claiming over 750,000 lives, primarily among the elderly (Johns Hopkins University and Medicine, 2020; Xu and Li, 2020; Zhou et al., 2020). SARS-CoV-2 is the third coronavirus, after SARS-CoV and MERS-CoV, causing severe pneumonia in humans (Corman et al., 2018).

Emerging clinical and molecular biology data from COVID-19 patients have detected SARS-CoV-2 nucleic acids primarily in bronchoalveolar lavage fluid, sputum, and nasal swabs and less frequently in fibrobronchoscope brush biopsies, pharyngeal swabs, and feces; and with even lower positive rates in blood and urine (Ling et al., 2020; Puelles et al., 2020; Wang et al., 2020d; Wu et al., 2020b; Young et al., 2020; Zou et al., 2020a). Pathological investigations, including postmortem biopsies, have confirmed major pulmonary damage as the most likely cause of death in the cases examined (Bradley et al., 2020; Huang et al., 2020; Xu et al., 2020b). There is also evidence that SARS-CoV-2 infection can damage other organ systems, including the heart, kidney, liver, and gastrointestinal (GI) tract, as documented previously for SARS and MERS (Ding et al., 2003; Gu et al., 2005; Gupta et al., 2020; Ng et al., 2016). Notably, it has been reported that cardiac and acute kidney injury is common in COVID-19 patients (Braun et al., 2020; Cheng et al., 2020; Diao et al., 2020; Fanelli et al., 2020; Shi et al., 2020; Anti-2019-nCoV Volunteers et al., 2020; Wang et al., 2020c). Severe COVID-19 patients show frequent liver dysfunctions (Zhang et al., 2020), and GI infection has also been reported (Gao et al., 2020; Xiao et al., 2020). Evidence of impaired gonadal function in male COVID-19 patients was also recently presented (Ma et al., 2020; Wang et al., 2020c). Intriguingly, SARS-CoV-2 can also be detected in the brain or cerebrospinal fluid and may even cause neurological complications (Ellul et al., 2020; Moriguchi et al., 2020; Wu et al., 2020c). What causes the wide range of clinical phenotypes observed in people infected with SARS-CoV-2 is not yet understood. It remains unclear which of these pathologies are caused by direct infection of the organs affected or indirect effects mediated by systemic inflammatory responses or comorbidities. A prerequisite to resolve these questions is to gain a better understanding of the tropism of the virus; i.e., which tissues and cell types are permissive to SARS-CoV-2 infection.

Because SARS-CoV-2 is a novel virus, our current knowledge of cellular factors regulating its entry into cells is mostly derived from studies of SARS-CoV, MERS-CoV, and "commensal" human coronaviruses (hCoVs). The canonical entry mechanism of these coronaviruses is a two-step process mediated by the viral Spike (S) protein decorating the virion. First, the S protein must

bind directly to a cell-surface receptor; and second, it must be cleaved ("primed") by a cellular protease to enable membrane fusion. Thus, the cellular tropism of a coronavirus is conditioned not only by the expression of an adequate receptor on the cell surface but also by the presence of a host-encoded protease capable of cleaving the S protein, preferably at or close to the site of receptor binding (de Haan and Rottier, 2005; Tang et al., 2020). For both SARS-CoV and SARS-CoV-2, Angiotensin-Converting Enzyme 2 (ACE2) and Transmembrane Serine Protease 2 (TMPRSS2) have been identified as a prime receptor and a critical protease, respectively, for cell entry (Glowacka et al., 2011; Hoffmann et al., 2020; Li et al., 2003; Matsuyama et al., 2010; Wrapp et al., 2020). These findings have prompted numerous efforts to profile the basal expression levels of *ACE2* and/or *TMPRSS2* across healthy human tissues to predict the tropism of these two closely related viruses. While studies monitoring protein abundance *in situ* (e.g., immunocytochemistry) offer a more direct assessment and have been conducted previously to study ACE2 and/or TMPRSS2 expression (Hamming et al., 2004; Hikmet et al., 2020; Hoffmann et al., 2020), most recent investigations have taken advantage of single-cell RNA-sequencing (scRNA-seq) data to profile the expression of these two factors at cellular resolution in a wide array of tissues (Table S1).

Collectively these studies have revealed a subset of tissues and cell types potentially susceptible to SARS-CoV-2 (see Table S1). However, they suffer from several limitations. First, most studies (15/27) profiled a single organ or organ system, and the majority focused on the respiratory tract. Second, most studies (19/27) restricted their analysis to *ACE2* and/or *TMPRSS2*, ignoring other factors potentially limiting SARS-CoV-2 entry or replication. However, there is evidence that these two proteins alone cannot solely explain all the current clinical and research observations. For instance, certain cell lines (e.g., A549 alveolar lung carcinoma) can be infected by SARS-CoV-2 in the absence of appreciable level of *ACE2* RNA or protein (Blanco-Melo et al., 2020; Wyler et al., 2020). Similarly, clinical data point to SARS-CoV-2 infection of several organs, such as lung, bronchus, nasopharynx, esophagus, liver, and stomach, where *ACE2* expression could not be detected in healthy individuals (Hikmet et al., 2020; Zou et al., 2020b). Moreover, there are discordant reports as to where and how much *ACE2* may be expressed in certain cells, including alveolar type 2 (AT2) cells of the lung, which are widely regarded as a primary site of infection and tissue damage. Together, these observations suggest either that *ACE2* expression levels vary greatly between individuals or in the course of an infection (Ziegler et al., 2020) or that SARS-CoV-2 can use alternate receptor(s) to enter certain cell types. For instance, cell-surface protein Basignin (BSG, also known as CD147) has been shown to interact with the S protein *in vitro* and facilitate entry of SARS-CoV and SARS-CoV-2 in Vero and 293T cells (Vankadari and Wilce, 2020; Wang et al., 2020b). In fact, SARS-CoV and other hCoVs can utilize multiple cell-surface molecules to promote their entry into cells, including ANPEP (Yeager et al., 1992), CD209 (DC-SIGN) (Yang et al., 2004), CLEC4G (LSECtin) (Marzi et al., 2004), and CLEC4M (LSIGN/CD299) (Gramberg et al., 2005). Likewise, hCoVs can use a variety of cellular proteases to prime their S protein, in substitution for TMPRSS2 in a cell-type-specific manner. These include other members of the TMPRSS family (e.g., TMPRSS4) (Zang et al., 2020) but also cathepsins (CTSL/M) (Simmons et al., 2013) and FURIN (Millet and Whittaker, 2014; Walls et al., 2020). Just as importantly, previous studies have not taken into account the expression of host factors, such as LY6E (Pfaender et al., 2020) and IFITM proteins (Huang et al., 2011), known to oppose or restrict cellular entry of hCoVs, including SARS-CoV-2. Overall, our understanding of cellular factors underlying the potential tropism of SARS-CoV-2 remains very partial.

To begin addressing these gaps, we curated a list of 28 human genes referred to as SARS-CoV-2 and coronavirus-associated receptors and factors (SCARFs) (Figure 1A; Table S2) and surveyed their basal RNA expression levels across a wide range of healthy tissues. Specifically, we mined publicly available scRNA-seq datasets using consistent normalization procedures to integrate and compare the dynamics of SCARF expression in human preimplantation embryos (Yan et al., 2013), at the maternal-fetus interface (MFI) (Vento-Tormo et al., 2018), in male and female gonads (Sohni et al., 2019; Wagner et al., 2020), in 14 other adult tissues (Han et al., 2020), and in nasal brushing from young and old healthy donors (Deprez et al., 2020; Garcıá et al., 2019; Vieira Braga et al., 2019). Additionally, we use bulk transcriptomics for four organs of interest (lung, kidney, liver, and heart) from human, chimpanzee, and macaque (Blake et al., 2020) to examine the conservation of SCARF expression across primates.

RESULTS

SCARF Curation

Because many cellular factors involved in viral replication, such as those involved in transcription, translation, and other housekeeping functions, are unlikely to affect the tropism of the virus, we primarily focus on factors acting at the level of entry. Two factors have been established most rigorously to promote cellular entry of SARS-CoV-2 (and, previously, SARS-CoV) in human cells (Figure 1; Table S2): the ACE2 receptor and TMPRSS2 protease (Hoffmann et al., 2020). BSG is a putative alternate receptor for both SARS-CoV and SARS-CoV-2, which has received experimental support (Chen et al., 2005; Wang et al., 2020b). We also included receptors that have been confirmed experimentally to facilitate entry of either SARS-CoV/hCoV-229E (ANPEP, CD209, and CLEC4G/M) or MERS-CoV (DPP4) and are, therefore, candidates for promoting SARS-CoV-2 entry (Vankadari and Wilce, 2020). Next, we considered a number of cellular proteases, in addition to TMPRSS2, as alternative priming factors. TMPRSS4 was recently shown to be capable of performing this function for SARS-CoV-2 in human cells (Zang et al., 2020). TMPRSS11A/B has been shown to activate the S peptide of other coronaviruses (Kam et al., 2009; Zmora et al., 2018). Additionally, FURIN is known to activate MERS-CoV and, possibly, SARS-CoV-2 (Millet and Whittaker, 2014; Walls et al., 2020), and cathepsins (CTSL/B) can also substitute TMPRSS2 to prime SARS-CoV (Simmons et al., 2013). Importantly, we also enlisted several restriction factors (RFs) that are known to protect cells against entry of SARS-CoV-2 (LY6E) (Pfaender et al., 2020) or a broad range of enveloped RNA viruses,

CellPress
OPEN ACCESS

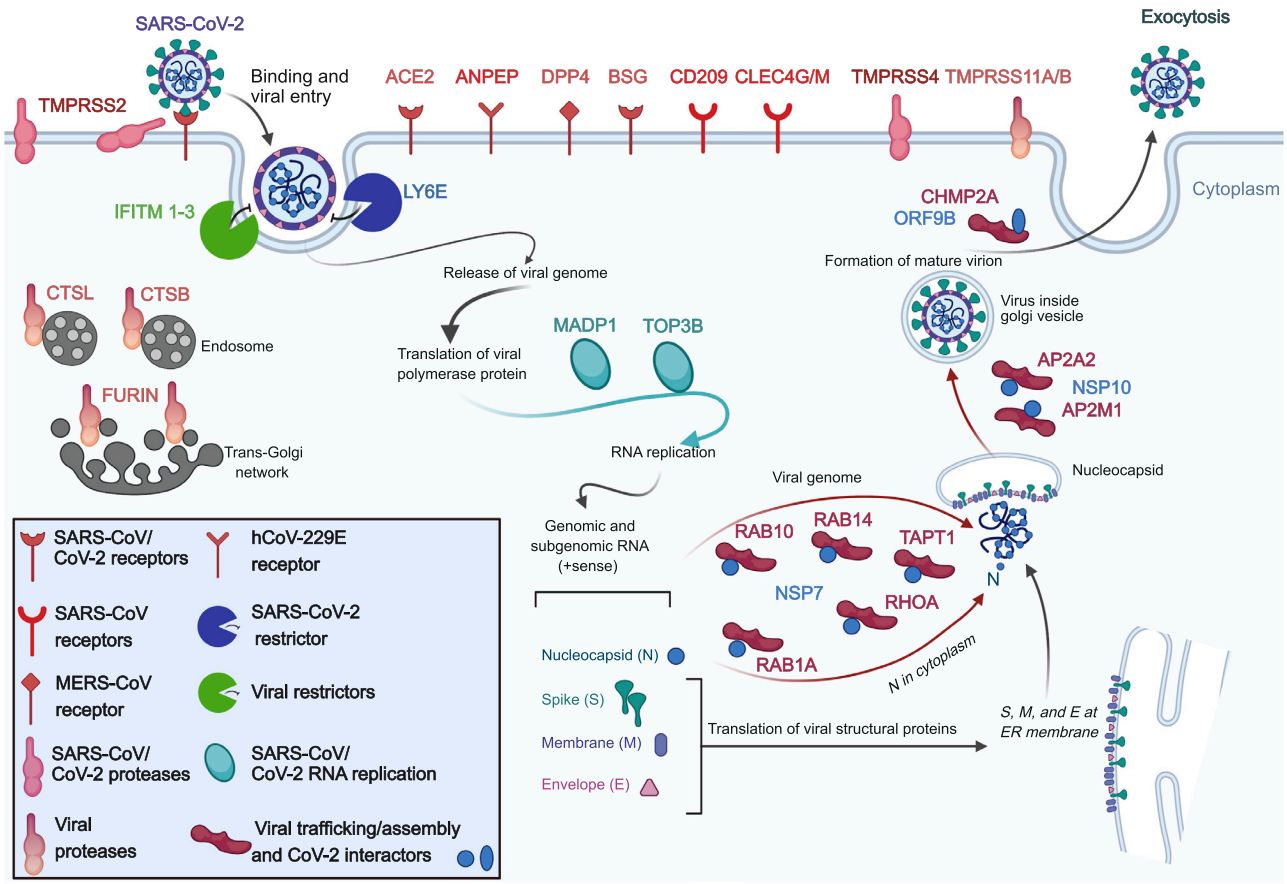


Figure 1. SARS-CoV-2 and SCARFs

A cartoon illustration of the infection cycle of SARS-CoV-2 and its cellular interaction with entry factors (cell surface receptors, proteases, and RFs) and post-entry factors (replication and assembly/trafficking factors) considered in this study.

including SARS-CoV (IFITM1-3) (Huang et al., 2011). We also considered a few additional factors that act post-entry but relatively early in the viral replication, such as TOP3B and MADP1 (ZCRB1), which may express in a tissue/cell-type-specific fashion and are known to be essential for SARS-CoV-2 and SARS-CoV genome replication, respectively (Prasanth et al., 2020; Tan et al., 2012). Lastly, we included a set of proteins known to be involved in assembly and trafficking of a range of RNA viruses and have been shown recently to interact physically with SARS-CoV-2 structural proteins (Gordon et al., 2020), including members of the Rho-GTPase complex (RHOA, RAB10, RAB14, and RAB1A), AP2 complex (AP2A2 and AP2M1), and CHMP2A. In total, our list includes 28 SCARFs we deem to be solid candidates for modulating SARS-CoV-2 entry and replication in human cells (Figure 1).

SCARF Expression during Preimplantation Embryonic Development

To profile SCARF RNA expression in early embryonic development, we mined scRNA-seq data for human preimplantation embryos (Yan et al., 2013). Our analysis revealed that *ACE2* is most abundant in the earlier stages of development, prior to zygotic genome activation (ZGA; 8-cell stage), indicating maternal RNA deposition (Figures 2A and S1A). *ACE2* levels are depleted post-ZGA until the trophectoderm (TE) formation, in which they rise up again (Figures 2A and S1A). By contrast, *TMPRSS2* expression is only apparent in TE lineages. In fact, none of the *TMPRSS* family members showed significant transcript levels ($\log_2$ FPKM > 1) prior to ZGA (Figures 2A, S1A, and S1B). Pluripotent stem cells show high expression of *IFITM1-3* but no evidence of *ACE2* expression (Figures 2A and S1A). Furthermore, analysis of another scRNA-seq dataset profiling ~60,000 cells representing cell types differentiated *in vitro* from pluripotent stem cells (Han et al., 2020) show no significant *ACE2* expression in any cells up to 20 days post-differentiation (Figure S1C). Together, these data suggest that pluripotent stem cells and cells in early stages of differentiation are unlikely to be permissive to SARS-CoV-2 infection.

SCARF Expression at the MFI

The high levels of *TMPRSS2*, *ACE2*, and other coronavirus receptors such as *ANPEP* in the TE, which gives rise to the

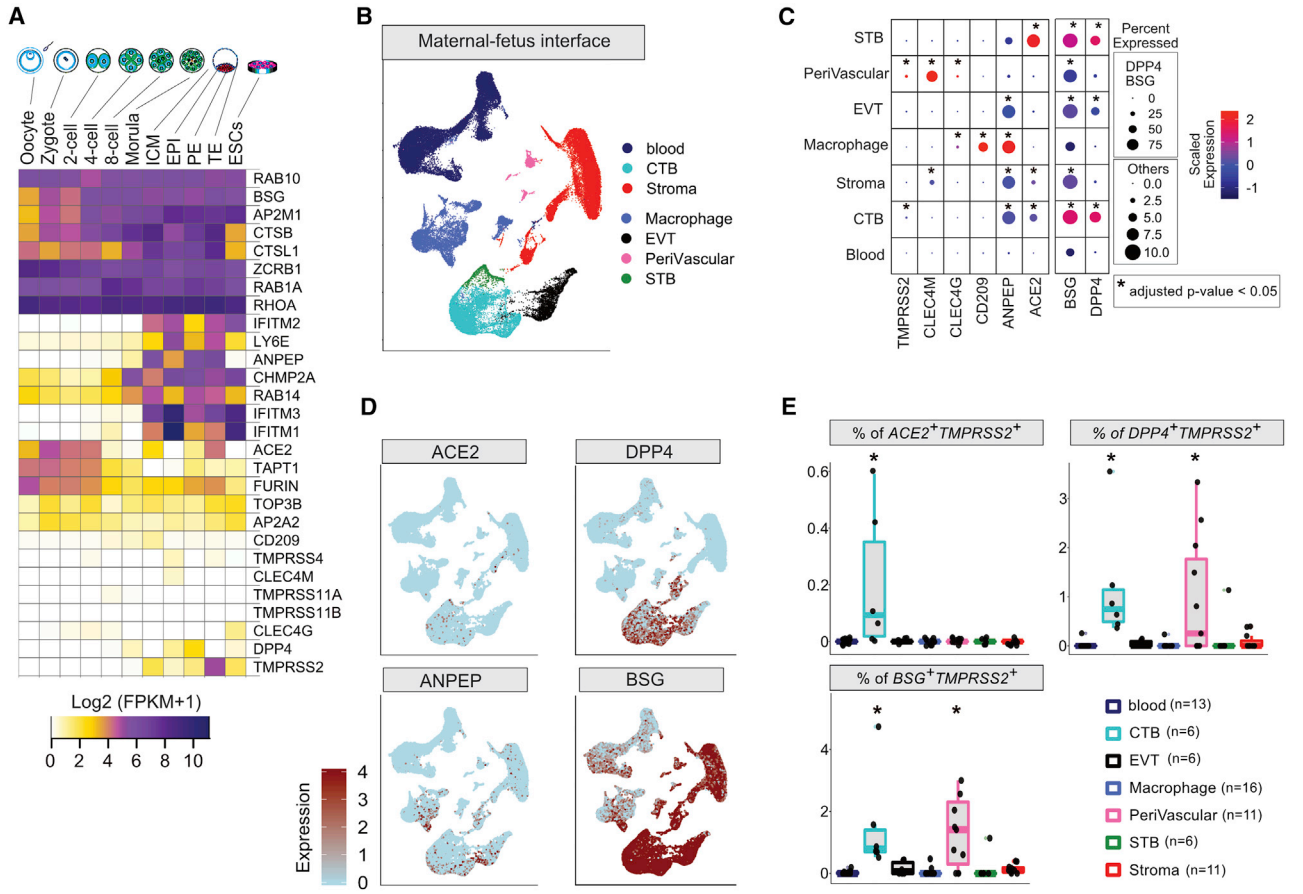


Figure 2. SCARF Expression in Preimplantation Embryo and at the MFI

(A) Heatmap of SCARF transcript levels in each stage of human preimplantation development (n = 20 oocytes and embryos, 124 single cells). RPKM, reads per kilobase of transcript per million mapped reads; ICM, inner cell mass; EPI, epiblast; PE, primitive endoderm; TE, trophectoderm; ESC, embryonic stem cell.

(B) Uniform Manifold Approximation and Projection (UMAP) plot illustrating cell clusters identified at the MFI (n = 22 donors, 64,054 single cells).

(C) RNA transcript intensity and density of SCARFs across the major cell types of the MFI. The dot color scales from blue to red, corresponding to lower and higher expression, respectively. The size of the dot is directly proportional to the percentage of cells expressing the gene in a given cell type. Star represents the significant enrichment of the percentage of positive cells calculated using one-sided Fisher's exact test and adjusted by Bonferroni correction.

(D) Feature plots displaying unsupervised identification of cells expressing ACE2, DPP4, BSG, and ANPEP over the UMAP of the MFI. Cells are colored according to expression levels, from light blue (low expression) to red (high expression).

(E) Boxplots showing the percentages of $ACE2^{+}TMPRSS2^{+}$, $DPP4^{+}TMPRSS2^{+}$, and $BSG^{+}TMPRSS2^{+}$ cells in the major cell types of the MFI (see also Table S4). Each dot represents an individual sample. Stars represent the adjusted p value obtained by Fisher's exact test adjusted by Bonferroni correction.

placenta, combined with low levels of *IFITMs* in this lineage (Figures 2A and S1A) raise the possibility that the developing placenta may be vulnerable to SARS-CoV-2 infection. To investigate this, we turned to the transcriptomes of ~70,000 single cells derived from tissues collected at the MFI during the first trimester of pregnancy (Vento-Tormo et al., 2018), which include both embryo-derived cells (placenta) and maternal blood and decidual cells. Our analysis of this dataset based on unsupervised clustering and examination of known markers recapitulated the major types of trophoblasts, decidua, and immune cells (Figures 2B and S2A). Expression of *ACE2* and *DPP4* receptors was evident in cytotrophoblasts (CTBs) and syncytiotrophoblasts (STBs). *ANPEP* was abundantly expressed in all fetal lineages, while BSG was broadly expressed in maternal and fetal cells but at a higher density in fetal cells (Figures 2C and 2D). Our analysis identified the *CLEC4M* gene as a strong marker of decidual perivascular cells (Figure 2C). Interestingly, extravillous trophoblasts (EVTs) showed low levels of *ACE2* or *TMPRSS2* but moderate to high levels of RFs *IFITM1-3* and *LY6E*, which were also expressed by immune and decidual cells (Figure S2B). *TMPRSS2*-expressing cells were comparatively less abundant within any cell types than those expressing receptors (Figure 2C). Thus, the maternal-placenta interface displays a complex pattern of SCARF expression.

To more finely assess the permissiveness of different placental cell types to SARS-CoV-2 entry, we quantified the fraction of each cell type co-expressing different combinations of receptors with proteases (predicted as more permissive) or with RFs (less permissive). The CTBs stood out for having the largest fraction of cells double-positive for various receptor-protease

CellPress
OPEN ACCESS

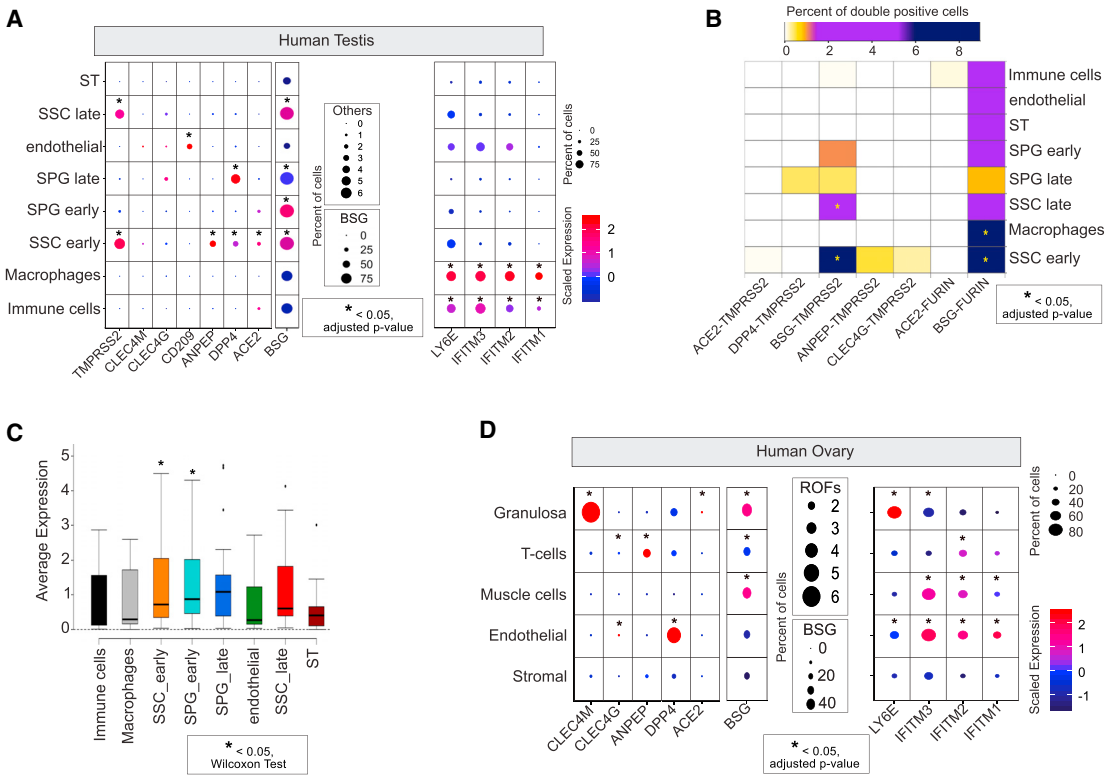


Figure 3. SCARF Expression in Reproductive Organs

(A) SCARF expression in the adult testis (n = 2, 6,572 single cells). ST, spermatid; SPG, spermatogonia; SSC, spermatogonial stem cell.

(B) Heatmap of the fraction of double-positive cells for different receptor-protease combinations in each of the major cell types of the testis. Color scheme includes white for the lowest fraction (<0.05% of cells), gold for medium (0.05%–2%), purple for high (2%–6%), and dark blue for the highest fraction (>6%) of double-positive cells per cell type.

(C) Boxplots showing average single-cell expression of SCARFs in different testis cell types.

(D) SCARF expression in the adult ovary (n = 5, 12,160 single cells). Note that TMPRSS2 and TMPRSS4 transcripts were not detected (UMI = 0) in any single cell of the ovary.

combinations, including *ACE2*$^{+}$*TMPRSS2*$^{+}$ (0.05%), *ACE2*$^{+}$*FURIN*$^{+}$ (~3%), *BSG*$^{+}$*TMPRSS2*$^{+}$ (0.8%), *BSG*$^{+}$*FURIN*$^{+}$ (10%), *DPP4*$^{+}$*TMPRSS2*$^{+}$ (0.6%), and *DPP4*$^{+}$*FURIN*$^{+}$ (~10%) (Figures 2D, 2E, and S2C; Table S4). Perivascular tissues also exhibited *BSG*$^{+}$*TMPRSS2*$^{+}$ and *DPP4*$^{+}$*TMPRSS2*$^{+}$ cells, albeit in fewer proportion compared to CTB (0.5%) (Figures 2E and S2B). Interestingly, a substantial fraction of *DPP4*$^{+}$ cells (~20%–80%) were co-expressing *IFITM1–3* and *LY6E* consistently across the whole dataset, whereas *ACE2*$^{+}$ and *BSG*$^{+}$ cells rarely co-expressed these RFs (Figures 2E and S2D; Table S4). Rather, *ACE2*$^{+}$ cells tend to co-express *TMPRSS2* and *FURIN*, but again, this was mostly confined to a small subset of CTB cells (Figures 2E and S2C). Overall, these results suggest that the CTB is the cell type most susceptible to coronavirus infection within the first-trimester placenta.

SCARF Expression in Reproductive Organs

Of all adult tissues surveyed via bulk RNA-seq by the GTEx Consortium (Battle et al., 2017), *ACE2* showed the highest level of expression in human testis (Figure S4A). To monitor more finely the expression profile of SCARFs in male and female reproductive tissues, we analyzed scRNA-seq datasets from testis samples collected from two healthy donors (Sohni et al., 2019) and adult ovary samples from five healthy donors (Wagner et al., 2020). In adult testis, we were able to recapitulate the clusters identified in the original report (Sohni et al., 2019), which consisted of early and late stages of spermatogonia (SPGs), spermatogonial stem cells (SSCs), spermatids (STs), macrophages, endothelial cells, and immune cells, each defined by a unique set of marker genes (Figure S3A). Turning to SCARFs, we observed that *TMPRSS2* is strongly expressed in both early and late SSCs, whereas CoV receptors are abundantly expressed in the early stage of SSCs (Figure 3A). Early SSCs expressing one of the receptors (*ACE2*, *BSG*, *DPP4*, or *ANPEP*) were found to be consistently enriched for co-expression with *TMPRSS2* across the testis dataset (Figure 3B). Moreover, other SCARFs interacting with SARS-CoV-2 proteins and predicted to facilitate virus trafficking or assembly (Table S2) also show the highest transcript levels in SSCs and SPGs (Figure 3C). In contrast, RFs were lowly expressed in all four clusters of spermatogonial cells (Figure 3A). Overall, these observations indicate that spermatogonial cells may be highly permissive to SARS-CoV-2 infection.

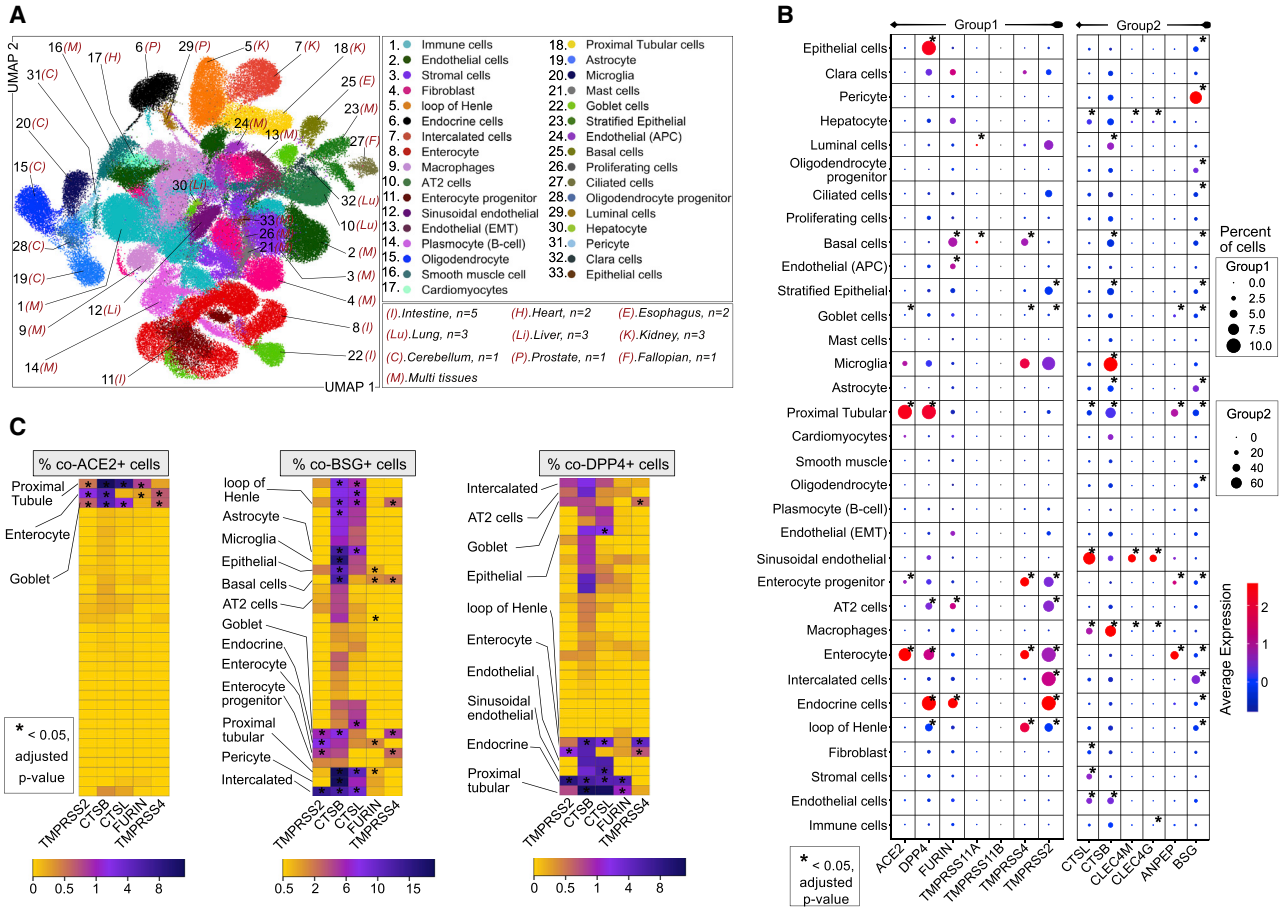


Figure 4. SCARF Expression across 14 Adult Tissues

(A) UMAP resolving ~200,000 single cells from 14 adult tissues obtained from the HCL into 33 different cell types. These cell types were consolidated from 78 cell clusters defined from the initial clustering (see also STAR Methods and Figure S5 for tissues corresponding to each cell type). Number of samples (n = 21 donors) is labeled on the plot.

(B) Expression intensity and density of selected SCARFs in each of the 33 cell types defined in Figure 4A. Different dot size scales are used for genes with a low fraction of positive cells (group 1: up to 10%) and those with higher fractions (group 2: up to 60%).

(C) Heatmap of the fraction of double-positive cells for different receptor-protease combinations in each of the 33 cell types.

Our analysis of ovarian cortex samples from 5 donors resolves the major cell types characteristic of this tissue, such as granulosa, immune, endothelial, perivascular, and stromal cells, each defined by a unique set of markers (Figure S3B), in concordance with the original article (Wagner et al., 2020). *ACE2*$^{+}$ cells were generally rare across this dataset and most evident in granulosa, where they show a relatively high level of expression per cell (Figure 3D). Alternate receptors were expressed at significant levels in specific ovarian cell populations. For instance, *DPP4* was expressed in 4% of endothelial cells, while *CLEC4M* and *BSG* were highly expressed in granulosa (Figure 3D). T cells and endothelial layers were markedly enriched for *ANPEP* and *DPP4* transcripts, respectively (Figure 3D). Strikingly, however, we could not identify any single cell across the entire ovarian dataset with evidence of *TMPRSS2* expression or any of the alternate proteases *TMPRSS4*, *TMPRSS11A*, and *TMPRSS11B*. This pattern was corroborated by examining the expression profile of all TMPRSS protease family members using bulk RNA-sequencing (RNA-seq) data from GTEx. None of these proteases appear to be expressed in ovarian tissue, whereas the testis expresses 7 out of 19 members, including *TMPRSS2* (Figure S4B). We note that the oocyte also lacks transcripts from any *TMPRSS* family members (see Figures 2A and S1A). Collectively, these data reveal a stark contrast between male and female reproductive organs: while early stages of spermatogenesis may be highly permissive for SARS-CoV-2 entry, the ovary and oocytes are unlikely to get infected.

Human Cell Landscape Reveals Adult Organs Potentially Most Permissive to SARS-CoV-2

We analyzed a scRNA-seq dataset for ~200,000 cells generated as part of the human cell landscape (HCL) project, which encompasses all major adult organs (Han et al., 2020). We selected 14 distinct adult tissues and clustered them into 33 distinct cell types annotated using the markers described in the original article (Han et al., 2020) (Figures 4A and S5A).

We first focused our analysis on *ACE2* and *TMPRSS2*, the two genes most robustly established as entry factors for SARS-CoV-2 (Table S2). Transcripts for both genes were detected (>0.1% of total cells) primarily in colon, intestine (ileum, duodenum, and jejunum), gallbladder, and kidney (Figures S5B–S5D) cells. We detected little to no expression of *ACE2* and/or *TMPRSS2* in the remaining tissues and cell types (Figure 4B), including alveolar type 1 (AT1) and AT2 cells of the lung. This latter observation is at odds with earlier reports (Table S1) but in line with a recent study using a wide array of techniques to monitor *ACE2* expression within the lung, including transcriptomics, proteomics, and immunostaining (Hikmet et al., 2020). It is also important to emphasize that even when *ACE2* and/or *TMPRSS2* were detected at appreciable RNA levels in a given organ or cell type, only a tiny fraction of cells was found to express both genes simultaneously. For instance, the kidney shows relatively high levels of *ACE2* and *TMPRSS2*, but only 0.01% cells were double-positive for these factors. Only three cell types stood out in our analysis for relatively elevated levels of co-expression of *ACE2* and *TMPRSS2* (0.5%–5% double-positive cells): enterocytes, proximal tubule cells, and goblet cells (Figures 4B and 4C), which we will return to later.

Examining alternative receptors revealed a more complex picture, in part reflecting the tissue specificity of these genes. For instance, *CLEC4G* and *CLEC4M* were highly expressed in the liver as previously reported (Domínguez-Soto et al., 2009) and, more specifically, in sinusoidal endothelial cells and hepatocytes (Figures 4B, S5B, and S5C). *BSG* marked the pericytes and astrocytes of the cerebellum, as well as intercalated cells of the kidney, whereas *CD209* was enriched in macrophages (Figures 4B and S5B). *ANPEP* and *DPP4* were often co-expressed in multiple organs and cell types, including prostate, lung, kidney, colon, and small intestine (Figure S5B). The prostate showed particularly high expression levels of *DPP4* (~15% double-positive with *TMPRSS2* or *FURIN*) and moderate levels of *ANPEP* (~5% double-positive with *TMPRSS2* or *FURIN*) (Table S5). Within the lung, AT2 cells also prominently expressed *DPP4*, *BSG*, and *ANPEP*, along with *TMPRSS2* and/or *FURIN* (Figures 4B and 4C), suggesting that these receptors, rather than *ACE2*, could represent the initial gateway by which SARS-CoV-2 infect AT2 cells, which are known to be extensively damaged in SARS and COVID-19 pathologies (Qi et al., 2020). Intercalated cells of the kidney as well as goblet cells and enterocytes of the colon were also highly enriched for *TMPRSS2*$^{+}$*DPP4*$^{+}$*BSG*$^{+}$ cells (Figures 4B, 4C, and S5A–S5C).

The small intestine was rather unique among the organs represented in this dataset for expressing high levels of *ACE2*, *ANPEP*, and *DPP4* (Figures S4A, S5B, and S5C), with highest levels in the jejunum, which also exhibited copious amount of *TMPRSS2* (Figures S5C–S5E and S6). Consistent with recent studies, we found that the bulk of expression of these factors in the small intestine is driven by enterocytes and their progenitors (Figures 4A–4C and S6) (Ziegler et al., 2020). Only two other cell types were equally remarkable for expressing the quadruple combination of *ACE2*, *ANPEP*, *DPP4*, and *TMPRSS2*: (1) goblet cells, epithelial cells lining and producing mucus for several organs including the duodenum, ileum, colon, and gallbladder; and (2) proximal tubular cells of the kidneys (Figures 4A–4C, S5A, S5B, and S6A–S6D). In addition, the enterocytes and goblet cells were enriched for SCARFs interacting with SARS-CoV-2 proteins and implicated in virus trafficking or assembly (Figure S6E).

In summary, coronavirus entry factors appear to be expressed in a wide range of healthy adult organs, but in restricted cell types, including pericytes, astrocytes, and microglia of the cerebellum, sinusoidal endothelium of the liver, endocrine cells of the prostate, enterocytes of the small intestine, goblet cells, and the proximal tubule of the kidney.

Nasal Epithelium

The nasal epithelium is thought to represent a major doorway to SARS-CoV-2 infection (Sungnak et al., 2020; Wu et al., 2020a). Since this tissue was not included in the HCL dataset, we analyzed another scRNA-seq dataset derived from six nasal brushing samples from healthy donors collected by three independent studies (Deprez et al., 2020; Garcıá et al., 2019; Vieira Braga et al., 2019) (Table S3). Our analysis of this merged dataset reveals four major cell clusters consistent across all 6 samples, corresponding to ciliated, secretory, and suprabasal epithelial cells and natural killer cells (Figures 5A and S7A). Natural killer cells express mostly *DPP4*, while the three epithelial cell types showed low to moderate amounts of *ACE2*, *ANPEP*, *BSG*, and *TMPRSS2* (Figure 5A; Table S6). *ACE2* was most abundant in ciliated cells, while *ANPEP* was most highly expressed in secretory cells. Conversely, *BSG* was abundant throughout all the nasal epithelial cell types, albeit at a higher level in suprabasal cells (Figures 5A and S7B; Table S6). *TMPRSS2* was also expressed in all three nasal epithelial cell types, with the highest density in ciliated cells (41%). In contrast with the digestive system (characterized earlier; Figures 4B and 4C), nasal epithelial cells rarely show co-expression of *ANPEP*, *DPP4*, and *ACE2* but rather exclusive expression of one of these three receptors (Figures 5A and S7B). By contrast, RFs *IFITM3* and *LY6E* were robustly expressed in all three nasal epithelial cell types (Figure 5A), with the highest levels in secretory and suprabasal cells (Table S6). Lastly, we calculated the percentage of double/triple-positive cells for various combinations of *ACE2*, *TMPRSS2*, and RFs (Table S6). We found that 85% and 65% of *ACE2*$^{+}$*TMPRSS2*$^{+}$ ciliated cells are also positive for *LY6E* and *IFITM3*, respectively (Table S6). Together, these data suggest that the nasal epithelium expresses various combinations of factors that, in principle, could facilitate SARS-CoV-2 infection, but it also expresses robust basal levels of RFs, which may act as a strong protective barrier in this tissue.

Age May Modulate SCARF Expression in the Nasal Epithelium

To investigate a possible age effect in the expression of CoV entry factors within the nasal epithelium, we took advantage of the fact that three of the samples analyzed as described earlier were collected from relatively young donors (24–30 years old), while the other three came from older individuals (50–59 years old) (Figure S7A; Table S3). While this is a small sample, this enabled us to split the data into ''young'' and ''old'' groups and compare the percentages of secretory and ciliated cells positive for entry

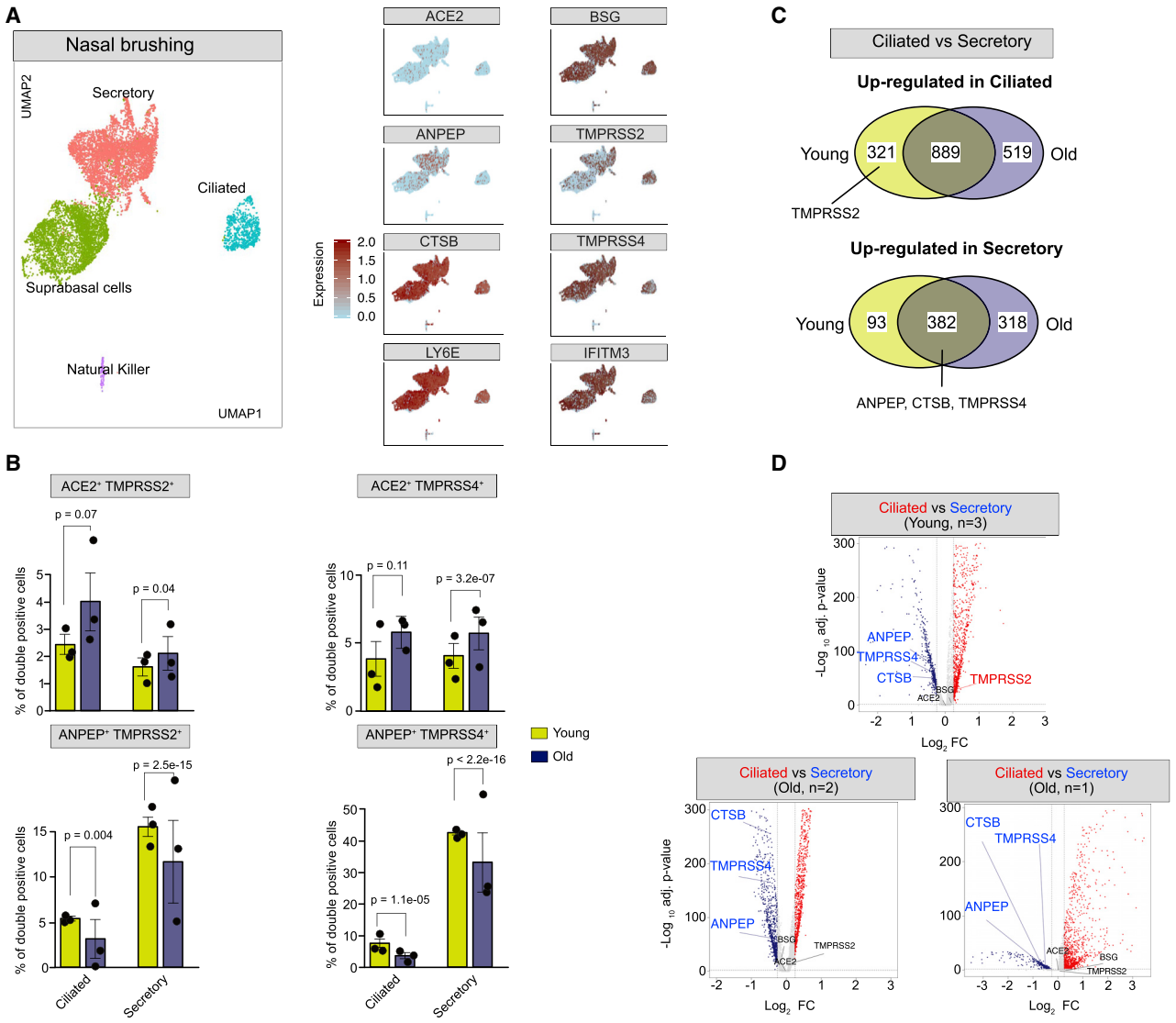


Figure 5. SCARF Expression in Nasal Brushing Samples

(A) Left panel: UMAP shows the cell clustering of 6 nasal epithelial scRNA-seq samples from 3 independent studies (n = 18,227 single cells). Right panel: feature plots showing expression of hCoV receptors (ACE2, ANPEP, and BSG), proteases (TMPRSS2, CTSB, and TMPRSS4), and RFs (LY6E and IFITM3).

(B) Bar plots comparing the percent of double-positive cells for different receptor/protease combinations in ciliated and secretory cells of the young group (n = 3) and old group (n = 3) nasal samples. Error bars denote the standard error from mean value. Each dot represents an individual sample. Data are represented as mean ± SEM.

(C) Venn diagrams illustrating shared and unique differentially expressed genes (DEGs) between ciliated and secretory cells within young group (n = 3) and old group (n = 3) samples. The upper panel indicates genes upregulated in ciliated cells, while the lower panel indicates genes upregulated in secretory cells.

(D) Volcano plots showing DEGs between ciliated and secretory cells identified independently for each of the three studies.

factors between the two groups (Table S3). The percentage of cells positive for *ACE2*, *TMPRSS2*, or *TMPRSS4* was comparable between the two groups, and these factors were most highly expressed in ciliated cells of both groups. However, the percentage of double-positive cells (*ACE2*$^{+}$*TMPRSS2*$^{+}$ or *ACE2*$^{+}$*TMPRSS4*$^{+}$) was significantly higher in the old group, within both ciliated and secretory cell populations (Figures 5B and S7C; Table S6). Interestingly, the percentage of *ANPEP*$^{+}$*TMPRSS2*$^{+}$/*4*$^{+}$ double-positive cells showed the opposite trend: they were significantly more frequent in the young group, within both ciliated and secretory cells (Figure 5B). To examine whether these differences were driven by an age-dependent shift in the relative expression of receptors and/or proteases within the nasal epithelium, we conducted a global differential gene expression analysis between ciliated and secretory cells within each age group and each independent study. We found that *ANPEP*, *TMPRSS4*, and *CTSB* were significantly upregulated in secretory cells relative to ciliated cells in all three

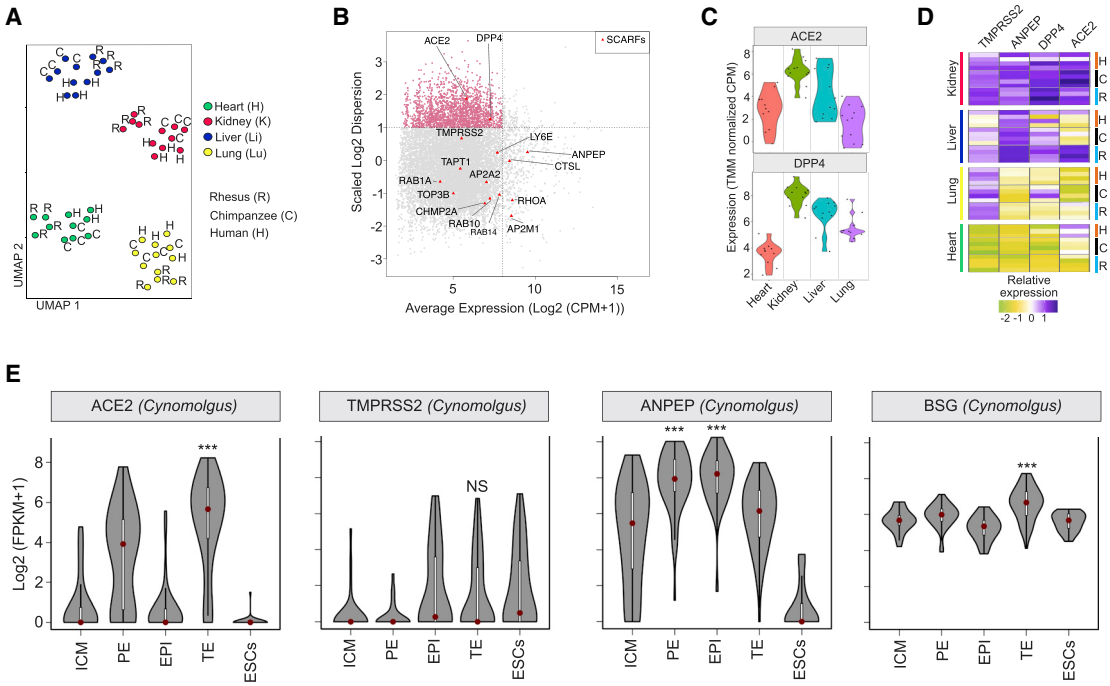


Figure 6. Cross-species Analysis of SCARF Expression

(A) UMAP clustering of 47 individual RNA-seq samples derived from four tissues (heart, lung, liver, and kidney) of three species (human, chimpanzee, and rhesus macaque) based on 11,929 orthologous RefSeq genes that were most variable in expression across the samples (see STAR Methods).

(B) Scatterplot of normalized mean expression in CPM (counts per million) of each orthologous gene (x axis) and scaled dispersion (y axis) across the 47 samples. Every point corresponds to a single gene. The most 1,987 variable genes across the samples ($\log_2$ scaled dispersion > 1) are indicated as pink dots. Red triangles indicate SCARFs.

(C) Violin plot showing normalized expression of ACE2 and DPP4 per tissue. Each dot comes from a different sample.

(D) Heatmap of scaled expression levels (TMM [trimmed mean of M]-normalized CPM) of ACE2, TMPRSS2, ANPEP, and DPP4 for each of the individual samples (n = 47) grouped by tissues and species.

(E) Violin plots showing $\log_2$-normalized expression profiles of ACE2, TMPRSS2, and ANPEP in the preimplantation blastocyst of cynomolgus macaque obtained by scRNA-seq.

studies, regardless of age (Figure 5C). Conversely, *TMPRSS2* levels were upregulated in ciliated cells of the young group but remained unaltered in individuals in the old group, regardless of the study of origin (Figures 5C and 5D; Table S7). These results suggest that there is a shift in *TMPRSS2* regulation during nasal epithelium differentiation in young individuals that is not occurring in old individuals (Figures 5C and 5D; Table S7).

Conservation of SCARF Expression across Primates

To examine the evolutionary conservation of SCARF expression across primates, we analyzed a comparative transcriptome dataset (bulk RNA-seq) of heart, kidney, liver, and lung primary tissues from humans (n = 4), chimpanzees (n = 4), and rhesus macaques (n = 4), comprising a total of 47 samples (Blake et al., 2020). As expected, Uniform Manifold Approximation and Projection (UMAP) analysis showed that the samples clustered first by tissue and then by species (Figure 6A).

Overall, we observed a low level of tissue specificity and a high level of inter-specific conservation in the expression of most SCARFs, with the notable exceptions of *ACE2* and *DPP4* receptors, which were included among the most variably expressed genes across tissues and species (Figure 6B). Specifically, liver and lung samples from chimpanzees showed relatively high *DPP4* expression but very low *ACE2* expression relative to humans and macaques (Figures 6C and 6D). Liver and lung samples from humans and chimpanzees did not express *ACE2*; however, the macaque liver showed high expression levels of *ACE2*, *DPP4*, *ANPEP*, and *TMPRSS2* genes (Figure 6D). These results suggest that macaques may be more prone to liver coronavirus infection than humans or chimpanzees. In agreement with our scRNA-seq analysis of the HCL dataset, we found that *ACE2*, *DPP4*, *ANPEP*, and *TMPRSS2* genes are all expressed at higher levels in the kidney of all three primates, relative to the other three organs. Thus, out of the four organs examined in this analysis, the kidney appears to be the most readily permissive to coronavirus infection in all three species (Figures 6C and 6D).

Finally, we analyzed the pattern of SCARF expression in the blastocyst of cynomolgus macaques (a close relative of the rhesus macaque) in comparison to that of the human blastocyst (Figures 2A and S1A). Intriguingly, the two species show distinct expression patterns for several entry factors. *TMPRRS2* was highly expressed in the human TE, but transcripts for this gene were essentially undetectable in the macaque blastocyst, including TE (Figure 6E). Also, *ANPEP* was downregulated in

the macaque TE compared with the rest of the blastocyst lineages, while it was upregulated in the human TE (compare Figure 6E with Figure S1A). Thus, there may be substantial differences in the susceptibility of human and macaque early embryos to coronavirus infection.

DISCUSSION

While it is clear that COVID-19 is primarily a respiratory disease that causes death via pneumonia, many unknowns remain as to the extent of tissues and cell types vulnerable to SARS-CoV-2. How host genetic factors interact with the virus and how they modulate the course of infection also remain poorly understood. Our study, along with several others (Table S1), have tapped into vast amounts of publicly available scRNA-seq data to profile the expression of host factors thought to be important for entry of SARS-CoV-2 in healthy tissues. Because the basal expression level of these factors determines, at least in part, the tropism of the virus, this information is foundational in predicting which tissues are more vulnerable to infection. These data are also important for guiding and prioritizing clinical interventions and pathological studies, including biopsies. Finally, this type of analysis has the potential to reveal possible routes of infection within and between individuals.

Our study distinguishes itself from all other studies reported thus far (Table S1) by the wider range of factors (SCARFs) examined across a large array of tissues. We interrogated a wide and unbiased set of organs, largely conditioned by the public availability of raw data (e.g., unique molecular identifier [UMI] counts), to apply uniform normalization procedures across datasets. For instance, we were able to integrate HCL samples that were prepared and processed through a single sequencing facility and platform. Also, it is important to emphasize that only two datasets examined here (placenta and nasal cavity from the Human Cell Atlas) were previously analyzed elsewhere (Sungnak et al., 2020). Thus, our analyses provide an independent replication of findings reported elsewhere for some tissues (see Table S1) as well as additional insights detailed later.

In this study, we examined the expression profile of a wide range of SCARFs (Table S2). In particular, we integrated the basal expression level of RFs, LY6E, and IFITMs. While these RFs are known to be interferon inducible (Jia et al., 2012; Mar et al., 2018), their basal level of expression is likely to be a key determinant of SARS-CoV-2 tropism. Indeed, it is well established that coronaviruses are equipped with multiple mechanisms to suppress interferon and other immune signaling pathways (Frieman and Baric, 2008; Gralinski and Baric, 2015), and there is growing evidence that SARS-CoV-2 infection also elicits a "muted" interferon response (Blanco-Melo et al., 2020; Wyler et al., 2020). As such, the basal level at which RFs are expressed in a given target cell may be the major obstacle the virus encounters at the onset of an infection. Nonetheless, it is difficult to infer from RNA levels alone whether the expression of these RFs will truly tip the balance toward viral resistance of a given tissue. It is even possible that some coronaviruses have adapted to use these RFs to their advantage, as proposed for hCoV-OC43 (Zhao et al., 2014). In the following text, we summarize and discuss how our results further extend our understanding of SARS-CoV-2 tropism and pathology.

The Developing Embryo May Be at Risk of Infection

ACE2 mRNA levels are very low throughout the early stages of embryonic development, including pluripotent stem cells. It remains silent in pluripotent stem cells in culture, even after 20 days of *in vitro* differentiation to various cell types. The idea that *ACE2* expression positively correlates with the level of cellular differentiation has been observed in the context of the airway epithelium (Jia et al., 2005).

The only preimplantation lineage with substantial *ACE2* expression is the TE, where *TMPRSS2* is also highly expressed. During embryonic development, the TE gives rise to all the different types of trophoblasts (placental fetal cells). Accordingly, *ACE2* and *TMPRSS2* expression persists in a subset of trophoblasts at least up to the first trimester of pregnancy. We identified a small population of CTBs co-expressing *TMPRSS2* with *ACE2, BSG*, and/or *DPP4* but exhibiting very low levels of *IFITM* and *LY6E* RFs. Thus, a subset of trophoblast cells may be permissive to SARS-CoV-2 entry. Overall, our findings are consistent with current clinical data suggesting that vertical transmission of SARS-CoV-2 from infected mother to fetus is plausible but probably rare (Baud et al., 2020; Chen et al., 2020a; Cui et al., 2020; Hosier et al., 2020; Schoenmakers et al., 2020; Zeng et al., 2020). Future studies should be directed at examining SCARF expression at different stages of pregnancy and evaluating whether SARS-CoV-2 infection could compromise pregnancy.

Ovarian Cells May Be Resistant to SARS-CoV-2, but Spermatogonial Cells Seem Highly Permissive

While we found no evidence for expression of any TMPRSS genes considered in female reproductive tissues, our analysis suggests that the early stages of spermatogenesis are vulnerable to SARS-CoV-2 and, likely, other coronaviruses. Indeed, we observe that spermatogonial cells (and stem cells) express high levels of *ACE2*, *TMPRSS2*, *DPP4*, and *ANPEP* and low levels of *IFITM* and *LY6E* RFs. One study reported high levels of *ACE2* expression in spermatogonial, Sertoli, and Leydig cells (Wang and Xu, 2020) but did not investigate other SCARFs. While it is unknown whether SARS-CoV-2 or other coronaviruses can infiltrate testes, it is notable that postmortem autopsies of male patients infected by SARS-CoV revealed widespread germ cell destruction, few or no spermatozoa in seminiferous tubules, and other testicular abnormalities (Xu et al., 2006). These and other observations (Fanelli et al., 2020; Wang et al., 2020c), together with our finding that prostate endocrine cells also appear permissive for SARS-CoV-2, call for pathological examination of testes as well as investigation of reproductive functions in male COVID-19 patients.

Respiratory Tract: How Does SARS-CoV-2 Infect Lung Cells?

Because COVID-19 and SARS are primarily respiratory diseases, the lung and airway systems have been extensively profiled for *ACE2* and *TMPRSS2* expression (Table S2), two SCARFs believed to be primary determinant for SARS-CoV-2 tropism. Paradoxically, as a whole, healthy lung tissues show only modest

expression for *ACE2* and *TMPRSS2*, which is readily apparent in a number of expression databases and widely available resources such as GTEx (Figure S4). Nonetheless, a number of studies mining scRNA-seq data reported marked expression of *ACE2* and/or *TMPRSS2* in a specific lung cell population: the AT2 cells (Chow and Chen, 2020; Travaglini et al., 2019; Zhao et al., 2020). However, these observations have been challenged by more recent studies (Aguiar et al., 2020; Hikmet et al., 2020). Our results are consistent with the notion that basal *ACE2* RNA levels in lung cells, including AT2 cells, are very low. Importantly, we observed that AT2 cells do co-express the putative alternate receptors *BSG*, *ANPEP*, and/or *DPP4* along with *TMPRSS2* at appreciable frequencies (0.2%–0.7%). Taken together, these data raise the question whether ACE2 is the primary receptor by which SARS-CoV-2 initiates lung infection. A non-mutually exclusive explanation is that *ACE2* expression is widely variable in the lung due to genetic or environmental factors. Consistent with the latter, evidence is mounting that *ACE2* can be induced by interferon and other innate immune signaling (Ziegler et al., 2020).

CNS and Heart: Do the Clinical Symptoms Manifest CoV-2 Infection?

Some COVID-19 patients show neurological symptoms (De Felice et al., 2020) such as encephalitis, strokes, seizures, and loss of smell, but it remains unclear whether SARS-CoV-2 can actively infect the CNS. In one case diagnosed with encephalitis, SARS-CoV-2 RNA was detected in the cerebrospinal fluid (Moriguchi et al., 2020). Some have reported *ACE2* expression in various brain cells (Chen et al., 2020b). We found that the alternate receptor *BSG* was abundantly expressed in pericytes and astrocytes, but in those cell types, *TMPRSS2/4* was not expressed. However, *FURIN* and *CTSB* proteases were often co-expressed with *BSG* in these cells, suggesting potential alternate routes of viral entry in those cell types. Because pericytes are located in the vicinity of the blood-brain barrier, these cells may act as a gateway to CNS infection.

Whether SARS-CoV-2 can infect the heart is another open question. Severe heart damage and abnormal blood clotting has been reported in a substantial fraction of COVID-19 patients (Shi et al., 2020; Wang et al., 2020a). We and others (Litviňuková et al., 2020) found that *ACE2* is expressed in cardiomyocytes, but the same cell population does not appear to express *TMPRSS2/4*, so it remains unclear how the virus could infiltrate cardiomyocytes. Nonetheless, we observe that *FURIN* was co-expressed with *ACE2* in a very small fraction (<0.1%) of cardiomyocytes. While it is unclear whether SARS-CoV-2 can use FURIN to prime infection (Litviňuková et al., 2020), our findings suggest a possible path to heart infection.

Nasal Epithelium: Niche or Battleground?

Recently, Sungnak et al. (2020) showed that SARS-CoV-2 entry factors are highly expressed in secretory and ciliated cells of the nasal epithelium. In agreement, we found that the percentage of $ACE2^+$ or $TMPRSS2^+$ cells was higher among ciliated cells than among secretory or suprabasal cells. Conversely, we found that the percentage of $ANPEP^+$ cells was higher in secretory or suprabasal cells than in ciliated cells, while *BSG* was broadly expressed throughout the nasal epithelium. We also note that 19% of suprabasal cells expressed *TMPRSS2*. However, the percentage of $ACE2^+TMPRSS2^+$ cells remained rather low across the nasal epithelium, while *IFITM3* and *LY6E* RFs showed high RNA levels throughout this tissue. Collectively, these observations point to the nasal epithelium as an early battleground for SARS-CoV-2 infection, the outcome of which may be critical for the pathological development of COVID-19.

It is clear that COVID-19 causes more severe complications in patients with advanced chronological age. One study reported that, paradoxically, *TMPRSS2* expression levels tend to mildly decrease with age in human lung tissue (Chow and Chen, 2020). We found that neither *ACE2* nor *TMPRSS2/4* on their own were differentially expressed between young group and old group nasal epithelia, but we observed that the percentage of $ACE2^+TMPRSS2^+/4^+$ double-positive cells was greater in older donors, both within ciliated and secretory cells. Conversely, the percentage of $ANPEP^+TMPRSS2^+/4^+$ cells was significantly higher in younger donors. It is tempting to speculate that the opposite susceptibility of old and young people to SARS-CoV and SARS-CoV-2 may relate to the differential usage of ANPEP and ACE2 receptors by these two closely related coronaviruses. However, our analysis is limited by a small sample size and many possible confounders such as gender, smoking status, and other genetic and environmental factors, which could not be controlled for.

Digestive System: Infection Hotspot?

Consistent with several studies (Hikmet et al., 2020; Sungnak et al., 2020; Ziegler et al., 2020), we found that the small intestine is one of the "hottest" tissues for co-expression of *TMPRSS2* with *ACE2*, but also for *DPP4* and *ANPEP* as reported previously (Venkatakrishnan et al., 2020). Within the small intestine, we found that the jejunum is where highest expression of these factors is achieved. This is in slight deviation from Ziegler et al. (2020), who suggested that the ileum had the maximum expression of *ACE2*. Regardless of which section of the small intestine, both studies converge on the finding that expression of these factors is largely driven by enterocytes and their progenitors, which line the inner surface of the intestine and are therefore directly exposed to food and pathogens.

Goblet cells represent another cell type commonly found in the digestive system that we also predict to be permissive for SARS-CoV-2 entry. These are epithelial cells found in the airway, intestine, and colon that specialize in mucosal secretion. We found that goblet cells have some of the highest level of co-expression of *TMPRSS2* with one or several receptors, including *ACE2*, *ANPEP*, *DPP4*, and *CD147*/BSG. Goblet cells within the nasal epithelium have also been identified as potentially vulnerable to SARS-CoV-2 (Sungnak et al., 2020). Overall, our analysis of the digestive system is concordant with several other studies pointing at the lining of the GI tract as a common site of SARS-CoV-2 infection (Lamers et al., 2020). This could explain the digestive symptoms (e.g., diarrhea) presented in COVID-19 cases (Wang et al., 2020a) as well as the detection of viral shedding in feces (Xu et al., 2020a). If so, fecal-oral transmission of SARS-CoV-2 may be plausible, but it remains to be rigorously investigated.

Table 1. Prediction of Human Coronavirus Tropism at Numerous Confidence Levels

Tissue	Cell Type	SARS-CoV-2	SARS-CoV-1	MERS-CoV	hCoV-229E	hCoV-NL63
Brain	astrocytes	3	3			
	pericyte	4	4			
Esophagus	basal cells	4	4			
Lung	AT2 cells	3	3	6		
	Clara cells	2	2	6		
Intestine/gall bladder	goblet cells	6	6	6	6	3
	enterocytes	6	6	6	6	3
Heart	cardiomyocyte	5	5			
Liver	sinusoidal endothelium	3	4	5	5	
	hepatocyte		4		5	
Kidney	intercalated cells	4	4	6		3
	proximal tubular	6	6	6	6	3
Nasal	ciliated[a]	6	6			3
	secretory[a]	6	6		6	3
	suprabasal[a]	4	4			3
Testes	spermatogonial stem cells (early)	4	4	5	6	3
Prostate	endocrine cells	4	4	6	6	
Placenta	cytotrophoblasts	6	6	6	6	3
	perivascular cells	4	4	6	6	3

This table displays the numerous confidence levels of coronavirus tropism in human cells. 1, high expression of predicted receptor and protease; 2, high expression of predicted receptor and validated protease; 3, high expression of *in vitro* validated receptor and predicted proteases; 4, high expression of *in vitro* validated receptor and validated protease; 5, high expression of *in vivo* validated receptor and predicted protease; and 6, high expression of *in vivo* validated receptor and validated proteases.
[a]High expression of restriction factors, suggesting a battleground.

Predicting the Coronavirus Tropism in Human Cells

While we focused most of our discussion on SARS-CoV-2, our study also provides a valuable resource for exploring the tropism of other human coronaviruses, including SARS-CoV, MERS-CoV, hCoV-229E, and hCoV-NL63 (Table 1). For instance, CLEC4G/M is a potential candidate for mediating SARS-CoV infection of sinusoidal cells of the liver (Xu et al., 2020b). We found that specific cell types of the placenta, testis, intestine, and kidney co-express ACE2, ANPEP, and DPP4, and therefore seem permissive for multiple human coronaviruses, i.e., SARS-CoV, MERS-CoV, hCoV-229E and hCoV-NL63 entry (see Table 1). AT2 and Clara cells of the lungs appear particularly permissive to MERS-CoV, while the secretory cells of the nasal epithelium are likely susceptible to hCoV-229E infections (Table 1). While hCoV-NL63 is known to use ACE2 as a receptor, this virus may not rely on similar proteases as SARS-CoV/CoV-2 for activation (Esposito et al., 2006; Huang et al., 2006); thus we are less confident in predicting the tropism of hCoV-NL63 based on our analysis (Table 1). Lastly, the tropism of a few other human coronaviruses, such as hCoV-OC43 and hCoV-HKU1, could not be assessed in our study, because their entry is dependent on post-translational modification of host-encoded peptides (Hulswit et al., 2019).

Conclusions

Overall, this study provides a valuable resource for future studies of the basic biology of SARS-CoV-2 and other coronaviruses as well as clinical investigations of the pathology and treatment of COVID-19. We also established an open-access, web-based interface, dubbed SCARFace (https://cells.ucsc.edu/?ds=scarface), allowing any user to explore the expression of SCARFs (and any other human RefSeq gene) within any of the scRNA-seq dataset analyzed here. Our finding that SCARF expression is generally well conserved across primate species, along with the high level of sequence conservation of the ACE2 interface with the Spike protein (Damas et al., 2020; Melin et al., 2020), suggests that non-human primates are adequate models for the study of SARS-CoV-2 and the development of therapeutic interventions, including vaccines. Our study is limited by the constraints and shortcomings of scRNA-seq. These include the lack or under-representation of certain cell types that are rare or undetected due to low sequencing depth, isolation biases, or statistical cutoffs. Likewise, the expression level of any given gene may be underestimated due to dropout effects. Hence, we strongly recommend interpreting our results with caution, especially negative results such as ovarian cells lacking the TMPRSS2/4 expression. Furthermore, RNA expression levels are imprecisely reflective of protein abundance, and our observations need to be corroborated by approaches quantifying protein expression *in situ*. Lastly, and perhaps most critically, SCARF expression within and between individuals is bound to be heavily modulated by genetic and environmental factors, including infection by SARS-CoV-2 and other pathogens. Such variables may drastically shift the expression

patterns that we observe in healthy tissues from a limited number of donors. In fact, several SCARFs surveyed here such as IFITM and LY6E RFs (Jia et al., 2012; Mar et al., 2018), and apparently ACE2 itself (Ziegler et al., 2020), are known to be modulated by infection and the innate immune response. Our data still provide a valuable baseline to evaluate how SCARF expression may be altered during SARS-CoV-2 infection. Because we surveyed host factors associated with a range of zoonotic coronaviruses, this study may also prove a useful resource in the context of other eventual outbreaks.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Preimplantation embryos
  - Maternal-fetal interface (MFI)
  - Adult tissues
  - Nasal epithelium
  - Cross-species analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108175.

ACKNOWLEDGMENTS

We would like to thank members of the Feschotte Lab and of the Cornell Virology community and also Prof. Dr. Peter Heutink (Tübingen, Germany), Prof. Dr. Hannelore Ehrenreich (Göttingen, Germany), and Dr. Ankit Arora (Heidelberg, Germany) for helpful advice and discussions. M.S. is supported by a Presidential Postdoctoral Fellowship from Cornell University. V.B. is supported by a Career Development Fellowship at DZNE Tübingen. Work on host-virus interactions in the Feschotte lab is funded by R35 GM122550 from the National Institutes of Health. Figure 1 was created with BioRender.com.

AUTHOR CONTRIBUTIONS

C.F., M.S., and V.B. conceived the study. C.F. supervised the project and wrote the manuscript. M.S. and V.B. assisted in writing, carried out all data analysis, and constructed the web interface "*SCARFace*" (https://cells.ucsc.edu/?ds=scarface).

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 20, 2020
Revised: July 15, 2020
Accepted: August 31, 2020
Published: September 3, 2020

REFERENCES

Aguiar, J.A., Tremblay, B.J.-M., Mansfield, M.J., Woody, O., Lobb, B., Banerjee, A., Chandiramohan, A., Tiessen, N., Cao, Q., Dvorkin-Gheva, A., et al. (2020). Gene expression and *in situ* protein profiling of candidate SARS-CoV-2 receptors in human airway epithelial cells and lung tissue. Eur. Respir. J. Published online July 16, 2020. https://doi.org/10.1183/13993003.01123-2020.

Battle, A., Brown, C.D., Engelhardt, B.E., Montgomery, S.B., He, Y., Jo, B., Mohammadi, P., Park, Y.S., Parsana, P., Segrè, A.V., et al.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

Baud, D., Greub, G., Favre, G., Gengler, C., Jaton, K., Dubruc, E., and Pomar, L. (2020). Second-Trimester Miscarriage in a Pregnant Woman With SARS-CoV-2 Infection. JAMA *323*, 2198–2200.

Blake, L.E., Roux, J., Hernando-Herraez, I., Banovich, N.E., Perez, R.G., Hsiao, C.J., Eres, I., Cuevas, C., Marques-Bonet, T., and Gilad, Y. (2020). A comparison of gene expression and DNA methylation patterns across tissues and species. Genome Res. *30*, 250–262.

Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D., et al. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell *181*, 1036–1045.e9.

Bradley, B.T., Maioli, H., Johnston, R., Chaudhry, I., Fink, S.L., Xu, H., Najafian, B., Deutsch, G., Lacy, J.M., Williams, T., et al. (2020). Histopathology and ultrastructural findings of fatal COVID-19 infections in Washington State: a case series. Lancet *396*, 320–332.

Braun, F., Lütgehetmann, M., Pfefferle, S., Wong, M.N., Carsten, A., Lindenmeyer, M.T., Nörz, D., Heinrich, F., Meißner, K., Wichmann, D., et al. (2020). SARS-CoV-2 renal tropism associates with acute kidney injury. Lancet *396*, 597–598.

Breese, M.R., and Liu, Y. (2013). NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics *29*, 494–496.

Chen, Z., Mi, L., Xu, J., Yu, J., Wang, X., Jiang, J., Xing, J., Shang, P., Qian, A., Li, Y., et al. (2005). Function of HAb18G/CD147 in invasion of host cells by severe acute respiratory syndrome coronavirus. J. Infect. Dis. *191*, 755–760.

Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., Li, J., Zhao, D., Xu, D., Gong, Q., et al. (2020a). Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet *395*, 809–815.

Chen, R., Wang, K., Yu, J., Chen, Z., Wen, C., and Xu, Z. (2020b). The spatial and cell-type distribution of SARS-CoV-2 receptor ACE2 in human and mouse brain. bioRxiv. https://doi.org/10.1101/2020.04.07.030650.

Cheng, Y., Luo, R., Wang, K., Zhang, M., Wang, Z., Dong, L., Li, J., Yao, Y., Ge, S., and Xu, G. (2020). Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int. *97*, 829–838.

Chow, R.D., and Chen, S. (2020). The aging transcriptome and cellular landscape of the human lung in relation to SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2020.04.07.030684.

Corman, V.M., Muth, D., Niemeyer, D., and Drosten, C. (2018). Hosts and sources of endemic human coronaviruses. In Advances in Virus Research, *Vol. 100*, M. Kielian, T.C. Mettenleiter, and M.J. Roossinck, eds. (Elsevier), pp. 163–188.

Cui, P., Chen, Z., Wang, T., Dai, J., Zhang, J., Ding, T., Jiang, J., Liu, J., Zhang, C., Shan, W., et al. (2020). Clinical features and sexual transmission potential of SARS-CoV-2 infected female patients: a descriptive study in Wuhan, China. medRxiv. https://doi.org/10.1101/2020.02.26.20028225.

Damas, J., Hughes, G.M., Keough, K.C., Painter, C.A., Persky, N.S., Corbo, M., Hiller, M., Koepfli, K.-P., Pfenning, A.R., Zhao, H., et al. (2020). Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates. bioRxiv. https://doi.org/10.1101/2020.04.16.045302.

De Felice, F.G., Tovar-Moll, F., Moll, J., Munoz, D.P., and Ferreira, S.T. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and the Central Nervous System. Trends Neurosci. *43*, 355–357.

de Haan, C.A.M., and Rottier, P.J.M. (2005). Molecular interactions in the assembly of coronaviruses. Adv. Virus Res. *64*, 165–230.

Deprez, M., Zaragosi, L.-E., Truchi, M., Becavin, C., Ruiz García, S., Arguel, M.-J., Plaisant, M., Magnone, V., Lebrigand, K., Abelanet, S., et al. (2020). A Single-cell Atlas of the Human Healthy Airways. Am. J. Respir. Crit. Care Med. Published online July 29, 2020. https://doi.org/10.1164/rccm.201911-2199OC.

Diao, B., Feng, Z., Wang, C., Wang, H., Liu, L., Wang, C., Wang, R., Liu, Y., Liu, Y., Wang, G., et al. (2020). Human Kidney is a Target for Novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection. medRxiv. https://doi.org/10.1101/2020.03.04.20031120.

Ding, Y., Wang, H., Shen, H., Li, Z., Geng, J., Han, H., Cai, J., Li, X., Kang, W., Weng, D., et al. (2003). The clinical pathology of severe acute respiratory syndrome (SARS): a report from China. J. Pathol. *200*, 282–289.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Domínguez-Soto, A., Aragoneses-Fenoll, L., Gómez-Aguado, F., Corcuera, M.T., Clária, J., García-Monzón, C., Bustos, M., and Corbí, A.L. (2009). The pathogen receptor liver and lymph node sinusoidal endotelial cell C-type lectin is expressed in human Kupffer cells and regulated by PU.1. Hepatology *49*, 287–296.

Ellul, M.A., Benjamin, L., Singh, B., Lant, S., Michael, B.D., Easton, A., Kneen, R., Defres, S., Sejvar, J., and Solomon, T. (2020). Neurological associations of COVID-19. Lancet Neurol. *19*, 767–783.

Esposito, S., Bosis, S., Niesters, H.G.M., Tremolati, E., Begliatti, E., Rognoni, A., Tagliabue, C., Principi, N., and Osterhaus, A.D.M.E. (2006). Impact of human coronavirus infections in otherwise healthy children who attended an emergency department. J. Med. Virol. *78*, 1609–1615.

Fanelli, V., Fiorentino, M., Cantaluppi, V., Gesualdo, L., Stallone, G., Ronco, C., and Castellano, G. (2020). Acute kidney injury in SARS-CoV-2 infected patients. Crit. Care *24*, 155.

Frieman, M., and Baric, R. (2008). Mechanisms of severe acute respiratory syndrome pathogenesis and innate immunomodulation. Microbiol. Mol. Biol. Rev. *72*, 672–685.

Gao, Q.Y., Chen, Y.X., and Fang, J.Y. (2020). 2019 Novel coronavirus infection and gastrointestinal tract. J. Dig. Dis. *21*, 125–126.

Garcıá, S.R., Deprez, M., Lebrigand, K., Cavard, A., Paquet, A., Arguel, M.J., Magnone, V., Truchi, M., Caballero, I., Leroy, S., et al. (2019). Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. Development *146*, dev177428.

Glowacka, I., Bertram, S., Müller, M.A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T.S., He, Y., Gnirss, K., et al. (2011). Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J. Virol. *85*, 4122–4134.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468.

Gralinski, L.E., and Baric, R.S. (2015). Molecular pathology of emerging coronavirus infections. J. Pathol. *235*, 185–195.

Gramberg, T., Hofmann, H., Möller, P., Lalor, P.F., Marzi, A., Geier, M., Krumbiegel, M., Winkler, T., Kirchhoff, F., Adams, D.H., et al. (2005). LSECtin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. Virology *340*, 224–236.

Gu, J., Gong, E., Zhang, B., Zheng, J., Gao, Z., Zhong, Y., Zou, W., Zhan, J., Wang, S., Xie, Z., et al. (2005). Multiple organ infection and the pathogenesis of SARS. J. Exp. Med. *202*, 415–424.

Gupta, A., Madhavan, M.V., Sehgal, K., Nair, N., Mahajan, S., Sehrawat, T.S., Bikdeli, B., Ahluwalia, N., Ausiello, J.C., Wan, E.Y., et al. (2020). Extrapulmonary manifestations of COVID-19. Nat. Med. *26*, 1017–1032.

Hamming, I., Timens, W., Bulthuis, M.L.C., Lely, A.T., Navis, G., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J. Pathol. *203*, 631–637.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature *581*, 303–309.

Hikmet, F., Méar, L., Edvinsson, Å., Micke, P., Uhlén, M., and Lindskog, C. (2020). The protein expression profile of ACE2 in human tissues. Mol. Syst. Biol. *16*, e9610.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271–280.e8.

Hosier, H., Farhadian, S.F., Morotti, R.A., Deshmukh, U., Lu-Culligan, A., Campbell, K.H., Yasumoto, Y., Vogels, C.B., Casanovas-Massana, A., Vijayakumar, P., et al. (2020). SARS-CoV-2 infection of the placenta. J. Clin. Invest. *130*, 4947–4953.

Huang, I.C., Bosch, B.J., Li, F., Li, W., Lee, K.H., Ghiran, S., Vasilieva, N., Dermody, T.S., Harrison, S.C., Dormitzer, P.R., et al. (2006). SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. J. Biol. Chem. *281*, 3198–3203.

Huang, I.-C., Bailey, C.C., Weyer, J.L., Radoshitzky, S.R., Becker, M.M., Chiang, J.J., Brass, A.L., Ahmed, A.A., Chi, X., Dong, L., et al. (2011). Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. PLoS Pathog. *7*, e1001258.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506.

Hulswit, R.J.G., Lang, Y., Bakkers, M.J.G., Li, W., Li, Z., Schouten, A., Ophorst, B., van Kuppeveld, F.J.M., Boons, G.J., Bosch, B.J., et al. (2019). Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proc. Natl. Acad. Sci. USA *116*, 2681–2690.

Izsvák, Z., Wang, J., Singh, M., Mager, D.L., and Hurst, L.D. (2016). Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? BioEssays *38*, 109–117.

Jia, H.P., Look, D.C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C., Perlman, S., and McCray, P.B., Jr. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. J. Virol. *79*, 14614–14621.

Jia, R., Pan, Q., Ding, S., Rong, L., Liu, S.-L., Geng, Y., Qiao, W., and Liang, C. (2012). The N-terminal region of IFITM3 modulates its antiviral activity by regulating IFITM3 cellular localization. J. Virol. *86*, 13697–13707.

Johns Hopkins University and Medicine (2020). COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://coronavirus.jhu.edu/map.html.

Kam, Y.-W., Okumura, Y., Kido, H., Ng, L.F.P., Bruzzone, R., and Altmeyer, R. (2009). Cleavage of the SARS coronavirus spike glycoprotein by airway proteases enhances virus entry into human bronchial epithelial cells in vitro. PLoS ONE *4*, e7870.

Lamers, M.M., Beumer, J., van der Vaart, J., Knoops, K., Puschhof, J., Breugem, T.I., Ravelli, R.B.G., van Schayck, P., Mykytyn, A.Z., Duimel, H.Q., et al.

(2020). SARS-CoV-2 productively infects human gut enterocytes. Science *369*, 50–54.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature *426*, 450–454.

Ling, Y., Xu, S.B., Lin, Y.X., Tian, D., Zhu, Z.Q., Dai, F.H., Wu, F., Song, Z.G., Huang, W., Chen, J., et al. (2020). Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. Chin. Med. J. (Engl.) *133*, 1039–1043.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Fasouli, E.S., Samari, S., et al. (2020). Cells and gene expression programs in the adult human heart. bioRxiv. https://doi.org/10.1101/2020.04.03.024075.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.

Ma, L., Xie, W., Li, D., Shi, L., Mao, Y., Xiong, Y., Zhang, Y., and Zhang, M. (2020). Effect of SARS-CoV-2 infection upon male gonadal function: A single center-based study. medRxiv. https://doi.org/10.1101/2020.03.21.20037267.

Mar, K.B., Rinkenberger, N.R., Boys, I.N., Eitson, J.L., McDougal, M.B., Richardson, R.B., and Schoggins, J.W. (2018). LY6E mediates an evolutionarily conserved enhancement of virus infection by targeting a late entry step. Nat. Commun. *9*, 3603.

Marzi, A., Gramberg, T., Simmons, G., Möller, P., Rennekamp, A.J., Krumbiegel, M., Geier, M., Eisemann, J., Turza, N., Saunier, B., et al. (2004). DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. J. Virol. *78*, 12090–12095.

Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. J. Virol. *84*, 12658–12664.

Melin, A.D., Janiak, M.C., Marrone, F., Arora, P.S., and Higham, J.P. (2020). Comparative ACE2 variation and primate COVID-19 risk. bioRxiv. https://doi.org/10.1101/2020.04.09.034967.

Millet, J.K., and Whittaker, G.R. (2014). Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc. Natl. Acad. Sci. USA *111*, 15214–15219.

Moriguchi, T., Harii, N., Goto, J., Harada, D., Sugawara, H., Takamino, J., Ueno, M., Sakata, H., Kondo, K., Myose, N., et al. (2020). A first case of meningitis/encephalitis associated with SARS-Coronavirus-2. Int. J. Infect. Dis. *94*, 55–58.

Ng, D.L., Al Hosani, F., Keating, M.K., Gerber, S.I., Jones, T.L., Metcalfe, M.G., Tong, S., Tao, Y., Alami, N.N., Haynes, L.M., et al. (2016). Clinicopathologic, immunohistochemical, and ultrastructural findings of a fatal case of middle east respiratory syndrome coronavirus infection in the United Arab Emirates, April 2014. Am. J. Pathol. *186*, 652–658.

Paules, C.I., Marston, H.D., and Fauci, A.S. (2020). Coronavirus Infections – More Than Just the Common Cold. JAMA, Published online January 23, 2020. https://doi.org/10.1001/jama.2020.0757.

Pfaender, S., Mar, K.B., Michailidis, E., Kratzel, A., Boys, I.N., V'kovski, P., Fan, W., Kelly, J.N., Hirt, D., Ebert, N., et al. (2020). LY6E impairs coronavirus fusion and confers immune control of viral disease. Nat. Microbiol. Published July 23, 2020. https://doi.org/10.1038/s41564-020-0769-y.

Prasanth, K.R., Hirano, M., Fagg, W.S., McAnarney, E.T., Shan, C., Xie, X., Hage, A., Pietzsch, C.A., Bukreyev, A., Rajsbaum, R., et al. (2020). Topoisomerase III-β is required for efficient replication of positive-sense RNA viruses. Antiviral Res. *182*, 104874.

Puelles, V.G., Lütgehetmann, M., Lindenmeyer, M.T., Sperhake, J.P., Wong, M.N., Allweiss, L., Chilla, S., Heinemann, A., Wanner, N., Liu, S., et al. (2020). Multiorgan and Renal Tropism of SARS-CoV-2. N. Engl. J. Med. *383*, 590–592.

Qi, D., Yan, X., Tang, X., Peng, J., Yu, Q., Feng, L., Yuan, G., Zhang, A., Chen, Y., Yuan, J., et al. (2020). Epidemiological and clinical features of 2019-nCoV acute respiratory disease cases in Chongqing municipality, China: a retrospective, descriptive, multiple-center study. medRxiv. https://doi.org/10.1101/2020.03.01.20029397.

Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A., and Teichmann, S.A. (2017). The Human Cell Atlas: from vision to reality. Nature *550*, 451–453.

Schoenmakers, S., Snijder, P., Verdijk, R., Kuiken, T., Kamphuis, S., Koopman, L., Krasemann, T., Rousian, M., Broekhuizen, M., Steegers, E., et al. (2020). SARS-CoV-2 placental infection and inflammation leading to fetal distress and neonatal multi-organ failure in an asymptomatic woman. medRxiv. https://doi.org/10.1101/2020.06.08.20110437.

Shi, S., Qin, M., Shen, B., Cai, Y., Liu, T., Yang, F., Gong, W., Liu, X., Liang, J., Zhao, Q., et al. (2020). Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. JAMA Cardiol. *5*, 802–810.

Simmons, G., Zmora, P., Gierer, S., Heurich, A., and Pöhlmann, S. (2013). Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. Antiviral Res. *100*, 605–614.

Sohni, A., Tan, K., Song, H.W., Burow, D., de Rooij, D.G., Laurent, L., Hsieh, T.C., Rabah, R., Hammoud, S.S., Vicini, E., and Wilkinson, M.F. (2019). The Neonatal and Adult Human Testis Defined at the Single-Cell Level. Cell Rep. *26*, 1501–1517.e4.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., Talavera-López, C., Maatz, H., Reichart, D., Sampaziotis, F., et al.; HCA Lung Biological Network (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat. Med. *26*, 681–687.

Tan, Y.W., Hong, W., and Liu, D.X. (2012). Binding of the 5′-untranslated region of coronavirus RNA to zinc finger CCHC-type and RNA-binding motif 1 enhances viral replication and transcription. Nucleic Acids Res. *40*, 5065–5077.

Tang, T., Bidon, M., Jaimes, J.A., Whittaker, G.R., and Daniel, S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. Antiviral Res. *178*, 104792.

Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J., et al. (2019). A molecular cell atlas of the human lung from single cell RNA sequencing. bioRxiv. https://doi.org/10.1101/742320.

Vankadari, N., and Wilce, J.A. (2020). Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. Emerg. Microbes Infect. *9*, 601–604.

Venkatakrishnan, A.J., Puranik, A., Anand, A., Zemmour, D., Yao, X., Wu, X., Chilaka, R., Murakowski, D.K., Standish, K., Raghunathan, B., et al. (2020). Knowledge synthesis from 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. eLife *9*, e58040.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. Nature *563*, 347–353.

Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., et al. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med. *25*, 1153–1163.

Anti-2019-nCoV Volunteers; Li, Z., Wu, M., Yao, J., Guo, J., Liao, X., Song, S., Li, J., Duan, G., Zhou, Y., Wu, X., et al. (2020). Caution on Kidney Dysfunctions of 2019-nCoV Patients. medRxiv. https://doi.org/10.1101/2020.02.08.20021212.

Wagner, M., Yoshihara, M., Douagi, I., Damdimopoulos, A., Panula, S., Petropoulos, S., Lu, H., Pettersson, K., Palm, K., Katayama, S., et al. (2020). Single-

cell analysis of human ovarian cortex identifies distinct cell populations but no oogonial stem cells. Nat. Commun. *11*, 1147.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281–292.e6.

Wang, Z., and Xu, X. (2020). scRNA-seq Profiling of Human Testes Reveals the Presence of the ACE2 Receptor, A Target for SARS-CoV-2 Infection in Spermatogonia, Leydig and Sertoli Cells. Cells *9*, 920.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al. (2020a). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA *323*, 1061–1069.

Wang, K., Chen, W., Zhou, Y.-S., Lian, J.-Q., Zhang, Z., Du, P., Gong, L., Zhang, Y., Cui, H.-Y., Geng, J.-J., et al. (2020b). SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. bioRxiv. https://doi.org/10.1101/2020.03.14.988345.

Wang, S., Zhou, X., Zhang, T., and Wang, Z. (2020c). The need for urogenital tract monitoring in COVID-19. Nat. Rev. Urol. *17*, 314–315.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., and Tan, W. (2020d). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA *323*, 1843–1844.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. *8*, 281–291.e9.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260–1263.

Wu, C., Zheng, S., Chen, Y., and Zheng, M. (2020a). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV, in the nasal tissue. medRxiv,. https://doi.org/10.1101/2020.02.11.20022228.

Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., et al. (2020b). Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. Lancet Gastroenterol. Hepatol. *5*, 434–435.

Wu, Y., Xu, X., Chen, Z., Duan, J., Hashimoto, K., Yang, L., Liu, C., and Yang, C. (2020c). Nervous system involvement after infection with COVID-19 and other coronaviruses. Brain Behav. Immun. *87*, 18–22.

Wyler, E., Mösbauer, K., Franke, V., Diag, A., Gottula, L.T., Arsie, R., Klironomos, F., Koppstein, D., Ayoub, S., Buccitelli, C., et al. (2020). Bulk and single-cell gene expression profiling of SARS-CoV-2 infected human cell lines identifies molecular targets for therapeutic intervention. bioRxiv. https://doi.org/10.1101/2020.05.05.079194.

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X., and Shan, H. (2020). Evidence for Gastrointestinal Infection of SARS-CoV-2. Gastroenterology *158*, 1831–1833.e3.

Xu, S., and Li, Y. (2020). Beware of the second wave of COVID-19. Lancet *395*, 1321–1322.

Xu, J., Qi, L., Chi, X., Yang, J., Wei, X., Gong, E., Peh, S., and Gu, J. (2006). Orchitis: a complication of severe acute respiratory syndrome (SARS). Biol. Reprod. *74*, 410–416.

Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., et al. (2020a). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. Nat. Med. *26*, 502–505.

Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., et al. (2020b). Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir. Med. *8*, 420–422.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

Yang, Z.-Y., Huang, Y., Ganesh, L., Leung, K., Kong, W.-P., Schwartz, O., Subbarao, K., and Nabel, G.J. (2004). pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. J. Virol. *78*, 5642–5650.

Yeager, C.L., Ashmun, R.A., Williams, R.K., Cardellichio, C.B., Shapiro, L.H., Look, A.T., and Holmes, K.V. (1992). Human aminopeptidase N is a receptor for human coronavirus 229E. Nature *357*, 420–422.

Young, M.D., and Behjati, S. (2018). SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. bioRxiv. https://doi.org/10.1101/303727.

Young, B.E., Ong, S.W.X., Kalimuddin, S., Low, J.G., Tan, S.Y., Loh, J., Ng, O.T., Marimuthu, K., Ang, L.W., Mak, T.M., et al.; Singapore 2019 Novel Coronavirus Outbreak Research Team (2020). Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. JAMA *323*, 1488–1494.

Zang, R., Gomez Castro, M.F., McCune, B.T., Zeng, Q., Rothlauf, P.W., Sonnek, N.M., Liu, Z., Brulois, K.F., Wang, X., Greenberg, H.B., et al. (2020). TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. Sci. Immunol. *5*, eabc3582.

Zeng, L., Xia, S., Yuan, W., Yan, K., Xiao, F., Shao, J., and Zhou, W. (2020). Neonatal Early-Onset Infection With SARS-CoV-2 in 33 Neonates Born to Mothers With COVID-19 in Wuhan, China. JAMA Pediatr. *174*, 722–725.

Zhang, C., Shi, L., and Wang, F.S. (2020). Liver injury in COVID-19: management and challenges. Lancet Gastroenterol. Hepatol. *5*, 428–430.

Zhao, X., Guo, F., Liu, F., Cuconati, A., Chang, J., Block, T.M., and Guo, J.T. (2014). Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. Proc. Natl. Acad. Sci. USA *111*, 6756–6761.

Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2020.01.26.919985.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet *395*, 1054–1062.

Ziegler, C., Allon, S.J., Nyquist, S.K., Mbano, I., Miao, V.N., Cao, Y., Yousif, A.S., Bals, J., Hauser, B.M., Feldman, J., et al.; HCA Lung Biological Network (2020). SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues. Cell *181*, 1016–1035.e19.

Zmora, P., Hoffmann, M., Kollmus, H., Moldenhauer, A.S., Danov, O., Braun, A., Winkler, M., Schughart, K., and Pöhlmann, S. (2018). TMPRSS11A activates the influenza A virus hemagglutinin and the MERS coronavirus spike protein and is insensitive against blockade by HAI-1. J. Biol. Chem. *293*, 13863–13873.

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., et al. (2020a). SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N. Engl. J. Med. *382*, 1177–1179.

Zou, X., Chen, K., Zou, J., Han, P., Hao, J., and Han, Z. (2020b). Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Front. Med. *14*, 185–192.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
STAR v2.3.0	Dobin et al., 2013	https://github.com/alexdobin/STAR
Cell Ranger v3.1.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/count; RRID:SCR_017344
Seurat v3.1.1	Stuart et al., 2019	https://satijalab.org/seurat/; RRID:SCR_016341
Scrublet v0.2.1	Wolock et al., 2019	https://github.com/allonkleinlab/scrublet; RRID:SCR_018098
SoupX v1.2.2	Young and Behjati., 2018	https://github.com/constantAmateur/SoupX
Other		
Resource website *SCARFace*	This paper	https://cells.ucsc.edu/?ds=scarface
R scripts	This paper	https://github.com/bansalvi/COVID19-SCARFs

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Cedric Feschotte (cf458@cornell.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Links of the original/source data used in this paper are available in Table S3. An open-access, web-based interface, dubbed SCARFace, is available at https://cells.ucsc.edu/?ds=scarface. Code necessary to perform Fisher's exact test and calculate the number of positive cells is available at https://github.com/bansalvi/COVID19-SCARFs.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The current study did not produce any experimental data. Links of the original/source data used in this paper are available in Table S3.

METHOD DETAILS

Preimplantation embryos

Single-cell (sc) RNA-seq datasets from the preimplantation stages of development were downloaded in a raw format from ((Yan et al., 2013), GSE36552). RNaseq reads with MAP quality score < 30 were removed. Resulting reads were mapped to the human genome (hg19) using STAR (https://github.com/alexdobin/STAR) (Dobin et al., 2013) with defined settings, i.e.–alignIntronMin 20–alignIntronMax 1000000–chimSegmentMin 15–chimJunctionOverhangMin 15–outFilterMultimapNmax 20, and only uniquely mapped reads were considered for the calculation of expression. RPKM was calculated using bamutils (https://ngsutils.org/modules/bamutils/count/) (Breese and Liu, 2013) for individual genes annotated in the human RefSeq database. Note: Here, we used the expression matrix generated for a previously published work (Izsvák et al., 2016), using the pipeline as mentioned above.

Maternal-fetal interface (MFI)

We obtained the processed expression matrix (counts) from ((Vento-Tormo et al., 2018), E-MTAB-6701) for ~70,000 single cells representing the MFI. We then used Seurat (v3.1.1) (https://github.com/satija.lab/seurat) (Stuart et al., 2019) within the R environment (v3.6.0) for the processing the dataset. We kept the cells with minimum and maximum of 1,000 and 5,000 genes expressed ($\geq$1 count), respectively. Moreover, cells with more than 5% of counts on mitochondrial genes were filtered out. After filtering, 64,782 cells remained. The data normalization was achieved by scaling it with a factor of 10,000 followed by natural-log transformation.

Clustering was performed using the "FindClusters" function with default parameters, except the resolution was set to 0.1. We used the first 20 Principle Component (PC) dimensions in the construction of the shared-nearest neighbor (SNN) graph to generate 2-dimensional embeddings for data visualization using UMAP. Cell type assignment was performed based on the annotations provided by the original publication, albeit we grouped the clusters into broader lineages. T cell, B cell, Dendritic, NK-cells, and Monocytes were categorized into "blood," all decidual cells, except perivascular cells, were annotated as "stroma." Fetal lineages were grouped into the known groups as ExtravillousTrophoblast ("EVTB"), CytoTrophoblast ("CTB"), and Syncytiotrophoblast ("STB") cells. All the given annotations were further confirmed by their respective markers (Figure S2). The expression levels of each gene in a cluster corresponding to the average $\log_2$ expression level scaled to the number of unique molecular identifier (UMI) values captured in single cells.

Adult tissues

We mined the scRNA-seq of 2 tissue samples from adult Testis ((Sohni et al., 2019), GSE124263), 31 tissue samples from 5 ovaries ((Wagner et al., 2020), GSE118127), 29 samples of 14 adult tissues from Human Cell Landscape (HCL) ((Han et al., 2020), GSE134355) in the form of raw counts. We favored this dataset over the Human Cell Atlas project (Rozenblatt-Rosen et al., 2017) because the HCL samples were prepared uniformly and processed through the same sequencing platform and raw counts for unique molecular identifiers (UMI) were publicly available, which enabled adequate normalization using our scRNA-seq analysis pipeline. To avoid the cross-platform batch biasedness, we independently processed the samples taken from different studies. Datasets corresponding to the same tissue were merged into one before the downstream processing. We scaled and normalized the datasets using Seurat (v3.1.1) (https://github.com/satija.lab/seurat) within the R environment (v3.6.0) as described in the previous section. We used a similar pipeline for the downstream analysis, except for merging of the HCL samples. We fed the first 50 PCs as input to cluster and visualize the single cells using SNN graphs and UMAP methods. The top marker genes distinguishing the cell types were calculated using the "FindAllMarkers" function implemented in Seurat, (adjusted p value < 0.01 and log(fold-change) > 0.25) using a Wilcoxon Rank Sum test. We annotated the cell types using the markers obtained in this study and cross-referenced with the original article.

Nasal epithelium

To compare young and old nasal tissue cells, we defined samples with age ≤ 30 as young and with age ≥ 50 as old. We found no study that included healthy old and young nasal samples within the same experiment. Therefore, six samples were taken from three studies ((Deprez et al., 2020; Garcıá et al., 2019; Vieira Braga et al., 2019); Table S3). Note that cell annotations were used as provided by the original publications except for the sample D318 (a raw count matrix was available). Unless explicitly mentioned, broad annotation terms as ciliated and secretory have been used for the sake of consistency.

Processing of samples "4," "6," "D353," "D363" and "D367": Normalized counts and cell-type annotations provided by the original publications were used. "CellType" annotations with $\geq$ 100 cells were considered, giving "Ciliated_2" (n = 1,513), "Goblet_2" (n = 1,463) and "Goblet_1" (n = 4,017) for old samples "4" and "6," and "LT/NK" (n = 185), "Multiciliated_N" (n = 855), "Secretory_N" (n = 7,138) and "Suprabasal_N" (n = 1,640) for young samples "D353," "D363" and "D367."

Processing of sample D318 raw matrix: First, cellranger (v3.1.0) *reanalyze* was used to generate a filtered matrix of top 5,000 cells (https://github.com/10XGenomics/cellranger). Next, we used Scrublet (v0.2.1) (Wolock et al., 2019) for identifying doublets with an expected doublet rate of 0.03. Overall, 58 cells with scores higher than 0.2 were discarded. SoupX (v1.2.2) (Young and Behjati, 2018) was used to subtract the ambient RNA profiles from real expression values. Finally, we used Seurat (v3.1.1) within the R environment (v3.6.0) for filtering, normalization and cell-type identification for sample D318. The following data processing was performed: (1) Filtering. We kept the cells with a minimum and maximum of 1,000 and 5,000 genes expressed ($\geq$1 count), respectively. Moreover, cells with more than 10% of counts on mitochondrial genes were filtered out. After filtering, 2,987 cells remained. (2) Data normalization. Gene UMI counts for each cell were divided by the total number of counts in that cell and multiplied by 10,000. These values were then natural-log transformed. (3) Cell-type identification. The integration of sample D318 scRNA-seq data with remaining samples was performed using the top 2000 variable features. Clustering was performed using the "*FindClusters*" function with default parameters except resolution was set to 0.1 and the first 30 PCA dimensions were used in the construction of the shared-nearest neighbor (SNN) graph and to generate 2-dimensional embeddings for data visualization using UMAP. Cell types were assigned based on the annotations provided by the original publication of samples "D353," "D363" and "D367," giving "LT/NK" (n = 110), "Multiciliated_N" (n = 62), "Secretory_N" (n = 1,354) and "Suprabasal_N" (n = 1,461).

Differential proportion analysis: For each sample, the number of positive cells for a gene was calculated when the count was higher than 0. The percentage of positive cells for a gene in a cell type was compared against the percentage of positive cells for the gene in all the remaining cells.

The numbers were added separately for young and old samples. A percentage was calculated for each cell type. P value was estimated between the percentage of positive cells in overall young and old groups of samples using one-sided fisher's exact test and further adjusted for multiple hypotheses using Bonferroni correction. The same analysis was performed for all the other datasets used in this study. The percentage of positive cells for a gene in a cell type was compared against the percentage of positive cells for the gene in all the remaining cells. We calculated the overall percentage of positive cells (for a gene or gene pairs) by pooling all the samples together to reduce the effect of drop-out events in scRNA-seq.

Differential expression analysis: We used the "*FindAllMarkers*" function with default parameters, except the minimum percentage was set to 5%. Default cutoffs were used to identify significant DE genes with log FC of |0.25| and adjusted p value of less than 0.01. Genes below these cutoffs are shown in volcano plots for visualization purposes only. All detected genes are plotted in a two-dimensional graph with x axis representing the log-fold change in transcript levels (calculated with the 'Seurat' package) and the y axis representing the significance level ($-\log_{10}$ adjusted p value, Wilcoxon Rank Sum test, adjusted with Bonferroni correction). Higher value on the y axis denotes stronger significance levels.

Cross-species analysis

Trimmed mean of M (TMM) values normalized cross-species counts per million (CPM) values were imported in R, and variable features were identified using the "*FindVariableFeatures*" function implemented in the Seurat package using *mean.var.plot (mvp)* as a selection method. Clustering was performed using the "*FindClusters*" function with default parameters except the resolution was set to 1 and the first 10 PCA dimensions were used in the construction of the shared-nearest neighbor (SNN) graph and to generate 2-dimensional embeddings for data visualization using UMAP.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using the Seurat (v3.1.1) statistical package within R version 3.6.0 environment. The P value for differential proportion analysis was estimated between the percentage of positive cells using one-sided fisher's exact test and further adjusted for multiple hypotheses using Bonferroni correction. All statistical details, including the statistical tests, represented number of samples (n), dispersion and precision measures, can be found in the methods and/or legends of the respective figures when noted appropriately.